Evaluating Radio Frequency Interference Detection Algorithms for SMAP (Soil Moisture Active Passive)

Student: Omar Salama

Advisor: Dr. Joel Johnson

Department: Electrical & Computer Engineering

*'**Undergraduate Honors Research Distinction Thesis**'*

Acknowledgements

I would like to express my deepest appreciation and sincerity to my advisor for his support and guidance, Dr. Joel Johnson. As well as my educational institute, The Ohio State University, and their staff for their sponsorship and education.

I would like to thank Dr. Jeffrey Piepmeier, Dr. Priscilla Mohammed and the SMAP team for their incredible hospitality, coaching, and leadership received at NASA, Goddard.

I would like to acknowledge my family for the utmost sustenance and care one can ask for. Without you, this wouldn't have been possible.

ABSTRACT

SMAP (Soil Moisture Active Passive) is a mission to be launched by NASA to measure soil moisture of the Earth's land surface. The SMAP radiometer operates in the L-band protected spectrum (1400-1427 MHz) that is known to be vulnerable to radio frequency interference (RFI). Radiometric observations show substantial evidence of out-of-band emissions from neighboring transmitters and possibly illegally operating emitters. SMAP faces large levels of RFI and also significant amounts of low-level RFI equivalent to 0.1 K to 10 K of brightness temperature. Such low-level interference would be enough to jeopardize mission success without an aggressive mitigation solution.

A decision has been made to employ an advanced digital microwave radiometer, the first of its kind for spaceflight, for use on SMAP. The mission takes a multi-domain approach to RFI mitigation utilizing an innovative on-board digital detector backend with DSP algorithms to detect and filter out harmful interference. Four different baseline RFI detectors are run on the ground and their outputs combined for a maximum probability of detection to remove RFI within a footprint. The SMAP radiometer outputs the first four raw moments of the receiver system noise voltage in 16 frequency channels for measurement of noise temperature and kurtosis as well as complex cross-correlation products for measuring the third and fourth Stokes parameters.

Evaluating each of the four individual RFI detection algorithms is essential to ensure the highest efficiency produced by the maximum probability of detection. Receiver operating characteristic (ROC) curves are generated for each of the different detectors to evaluate performance. ROC curves plot the probability of detection versus false alarm rate. The optimum case would correspond to the highest probability of detection (PD) and lowest false alarm rate (FAR). A given threshold for the RFI algorithms would produce a corresponding (PD, FAR). The rest of

the line curve is graphed by varying threshold from a minimal value to a maximal value. The ROC curves are performed on all different RFI algorithm detectors which include time-domain, cross-frequency, kurtosis, and polarization detectors. Each detector operates differently and behaves differently under different injected RFI. Different injected RFI include pulsed and sinusoidal at different frequencies, amplitudes, and power.

The focus of the study is to optimize each of the given RFI detectors given any RFI signal. For example, since the cross-frequency algorithm uses only frequency resolution and no time resolution, its performance should be best for RFI that is localized in frequency. Since continuous wave (CW) RFI are localized in frequency by definition, as expected, the cross-frequency detector performed very well against CW RFI relative to other detectors. The RFI detection performance that is ultimately achieved will be a function of the threshold (that returns the highest PD versus lowest FAR), the nature of the RFI encountered, and radiometer system parameters such as the number of frequency channels and the integration period.

1. INTRODUCTION

The SMAP mission is a wide-scale mission containing multiple dimensions to the project. Thus, a brief overview of the different scopes underlying this study will be reviewed to ensure that all technical concepts are clear.

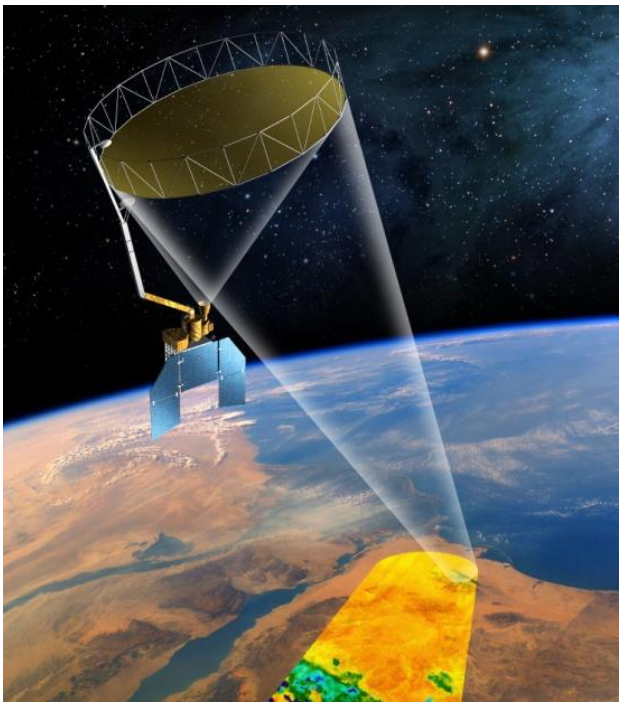


Figure 1: Computer Image of SMAP Satellite Orbiting Earth [18]

1.1 Satellite Scope & Operating Frequency

The SMAP satellite is designed to specifically monitor earth from orbit to measure temperature and soil moisture. The project satellite can be categorized as an earth observation satellite as ecological disciplines are studied through the data retrieved. The main mission is to provide global high-resolution active and passive L-band observations of the Earth from space. L-band is defined as the region of electromagnetic spectrum operating between 1-2 GHz. It is a protected band partly dedicated to satellite operation. The SMAP satellite it set to operate at a frequency

range between 1400-1427 MHz. The microwave radiometer's prime science objective is to provide soil moisture measurements with an uncertainty < 0.04 m^3m^{-3}, which requires a radiometric uncertainty < 1.3 K. This requires a highly accurate set of data, thus radio frequency interference cannot be tolerated.

1.2 Satellite Radiometer

A radiometer is a device utilized to measure radiant flux or power of electromagnetic radiation. An illustration of how a simple radiometer works could be presented through the "Crookes" radiometer. The outer shape of the "Crookes" radiometer looks like a light bulb. Inside are a set of vanes that are mounted on a spindle and are polished white on one side and black on the other. These vanes rotate at a speed proportional to the brightness of light they are exposed to. This leads to a qualitative measurement of electromagnetic intensity. However, more complicated radiometers are used to make scientific measurements of radiant energy. There are different types of radiometers, but the SMAP project will use a "microwave radiometer". Microwave radiometers are specialized to measure power of source frequencies of range [1-1000] GHz. The specific design and properties of the radiometer is determined after extensive studies have been put forth about band characterizations and RFI. The radiometer needs to be designed in a way to produce accurate and reliable data. A more detailed discussion of the SMAP radiometer used and it's specifics will be discussed in chapter 3.

1.3 Antenna & Wave Polarizations

An antenna may be defined as a conductor or group of conductors used either for radiating electromagnetic energy into space or for collecting it from space. Electrical energy from the transmitter is converted into electromagnetic energy by the antenna and radiated into space. On

the receiving end, electromagnetic energy is converted into electrical energy by the antenna and fed into the receiver. One property of an antenna is its polarization. The polarization of an antenna is the orientation of the electric field of the radio wave with respect to the Earth's surface and is determined by the physical structure of the antenna and by its orientation. SMAP has horizontally and vertically placed antennas for retrieving data. Furthermore, the radiometer is able to accept these polarizations as well as add a third and fourth dimension for further analysis. Stokes parameters are another set that describe the polarization of a wave. Horizontal and vertical polarizations are referred to as first and second stokes parameters. The third and fourth stokes parameters are computed by performing the complex correlation of the horizontal and vertical polarizations. The real part of the cross correlation corresponds to the third stokes, where the imaginary part corresponds to the fourth stokes.

1.4 RFI Characterization and Identification

Radio frequency or electromagnetic interference is disturbance caused by external sources to a current operating system. Sources for these interferences are often unidentified, but a wide range of sources must be taken into account for a full evaluation. Identified interferences are easier to characterize and thus help in the process of mitigation. Characterization and studies on the types of RFI that exist within satellite operations has been studied extensively for the SMAP radiometer. Understanding the scope of interference that the satellite might endure will be essential in developing a radiometer that will counter and remove that interference. The study has been conducted through three steps. First, identify interference from neighboring frequency bands based on known properties of their operating channels. Second, examine both ground and space flight data observations that have been conducted to characterize more sources of RFI.

Third, analyze previous data sets from similar types of previous missions such as SMOS (Soil Moisture and Ocean Salinity Satellite) and Aquarius.

1.5 Statistics and Probability Theory

Due to the high uncertainty in the interference of data, statistics and probability theory are applied to the radiometer products. An unknown outcome of an experiment is called a "random variable". Probability theory expresses information about expected outcomes averaged over many measurements. Each sampled set of data collected from the radiometer can be thought of as a random variable. The probability density function (pdf) provides a complete knowledge of a single random variable. It provides the likelihood for this particular random variable to occur at a given time. We will often refer in this document to first through fourth moments. These moments are characterized as mean, variance, skewness, and kurtosis respectively.

A. *First Moment: Mean*

Given the probability density of a function, it is possible to compute the mean or sometimes called "expected value". Mean is computed by taking each possible value of a random variable, x, multiplying it by its pdf, P(x) and summing results. The following can be illustrated with the equation:

$$E(x) = \mu = \sum x * P(x) \tag{1.1}$$

B. *Second Moment: Variance*

Variance is sometimes useful in understanding how results in a given measurement will fluctuate about this average. It is a measure of the average deviation of the random variable from its mean. Random variables with large standard deviations relative to the mean have outcomes that

fluctuate significantly in a given measurement. Variance can be represented through the following equation:

$$\sigma^2 = E[\big(x - E(x)\big)^2] \tag{1.2}$$

Note that the variance is represented by the symbol variance squared. The square root of variance is denoted the "standard deviation".

C. <u>Third Moment: Skewness</u>

Before defining skewness, let's define an nth central moment. The nth moment about the mean of a random variable, x, is the quantity:

$$u_n = E\big[\big(x - E(x)\big)^n\big] \tag{1.3}$$

Also, an nth standardized moment is defined as: $\frac{u_n}{\sigma_n}$. The skewness of a random variable can then be defined as the third standardized moment where it can be written as:

$$\gamma = \frac{\mu_3}{\gamma_3} = \frac{E\big[\big(x-E(x)\big)^3\big]}{\Big(E\big[\big(x-E(x)\big)^2\big]\Big)^{\frac{3}{2}}} \tag{1.4}$$

Where, γ represents skewness. The qualitative interpretation of the skewness is complicated. However it can be thought of as giving information regarding the original PDF function shape. It can describe weather the random variable distribution is evened out over the center mean, more weighted to the right of the mean, or more weighted to the left of the mean. Skewness can return positive, negative, or zero values which indicate right-weighted, left-weighted, or even-weighted distributions along the mean respectively.

D. Fourth Moment: Kurtosis

Kurtosis, or the fourth moment, is another measure of characterizing the shape of the pdf. It focuses more on the peakness, and also weight of the PDF tails. Kurtosis free of the subtraction scaling is often defined as fourth standardized moment. The definition is very similar to Skewness, except a subtracting factor is introduced. The latter relationship is defined as,

$$K = \frac{\mu_4}{\gamma_4} - 3 = \frac{E\left[\left(x - E(x)\right)^4\right]}{\left(E\left[\left(x - E(x)\right)^2\right]\right)^2} - 3 \tag{1.5}$$

All these moments are introduced because they are used in analyzing flight data. These moments can become useful due to their increased specialty in what they provide. As a parallel thinking process, these can be thought of as accumulation of 'n' derivates where each 'n' derivate provides specific information about the original function. Here our original function can be thought of as the PDF, where each moment taken provides us more information about the original PDF function.

1.6 Temperature Brightness

A blackbody is a body that completely absorbs all wavelengths of thermal radiation incident on it. Such bodies do not reflect light, and therefore appear black if their temperatures are low enough so as not to be emitting light. The Planck law gives the intensity radiated by a blackbody as a function of frequency (or wavelength). Planck's law gives the intensity or brightness of a black body and can be defined as:

$$B_v(T) = I_v = \frac{2hv^3}{c^2} \frac{1}{e^{\frac{hv}{kT}} - 1} \tag{1.6}$$

Here h is Planck's constant; v is the frequency; c is the speed of light; k is Boltzmann's constant; T is the temperature of the black body; where I_v is the energy emitted per unit surface area per unit time per unit solid angle.

SMAP's final data measurements will be collected in units of brightness temperature. Brightness temperature is defined as the temperature of a black body that emits the same intensity as measured for the body. Brightness temperature is found by inverting Planck's law to achieve the relationship desired.

1.7 Calibration

The process of calibration can be thought of as converting one set of data to a referenced second set of data. The initial outputs of the radiometer require further processing algorithms to convert measured values to our final temperature brightness. Calibrating the radiometer outputs converts raw measured data into separate sets of data where special analysis and observations are formed. Calibration requires some ancillary information and references points to be added to the raw data in order to achieve an accurate and correct final temperature brightness output.

1.8 Approach and Steps

This document will proceed with different chapters describing the different scopes of the project in greater detail. The document will start with categorizing the types of RFI that is dealt with in Chapter 2. The following chapter will describe the radiometer logical block diagram flow as well as the raw outputs produced by the radiometer. Chapter 4 will discuss the processing and calibration algorithms used to convert the raw data into two levels of more data. It will also discuss the outputs of each level and the essential uses of these products for analysis. Chapter 5 will discuss all the baseline detection algorithms used for RFI mitigation. The final chapter gives

a description of receiver operating characteristic curves as which are key to evaluate these given baseline detection algorithms.

2. RFI ENVIRONMENT

Studies of the RFI environment have been conducted to improve the efficiency and focus of the RFI detection algorithms. The problem of RFI in spaceborne microwave radiometer measurements has been known for at least three decades. RFI was observed in the past at higher frequency bands like the C-band (4-8 GHz) in data from the Scanning Multichannel Microwave Radiometer (SMMR), which was being used for measuring sea surface temperature. More recently, data from AMSR-E and WindSat have confirmed that the X 97-11 GHz) and K (18-27 GHz) bands have been added to the list of ranges that include interference [1]. As the ranges of frequency decrease, it has been observed that interference becomes a higher reliability. Thus frequency range of SMAP, L-band at 1.4 GHz, has shown significant levels of RFI interference. While radio spectrum managers work within the radio regulations to enable compatibility between active and passive RF systems, there is no guarantee (even if all the rules are followed) that radiometer observations will be RFI free [7]. Fortunately, recent system designers are able to implement solutions to mitigate the deleterious effects of interference. The Aquarius radiometers on SAC-D and their ground processing software comprise the first such spaceborne system designed to detect and remove much of the interference caused by ground base radars [8-11]. More details regarding the range and scope of interference dealt with at the L-Band is described within.

2.1 Out-of-Band (OOB) Identification

The SMAP satellite operates at a protected L-band, thus it is highly unlikely to find illegal operating channels in the band. However, interference may still exist from legal sources operating outside the [1400- 1427] MHz spectrum. There are two main mechanisms by which these legally operating bands can cause RFI. In the first, the SMAP frequency response may be insufficient to eliminate the all the influence of radiation transmitted at frequencies outside the L band range. The SMAP filters are designed to have high rejection outside this range, thus the primary concern are for sources operating very near the 1400-1427 region. For example operating frequencies at 1395-1400 MHz would be a possible danger to our data. Second, it turns out that there are special emissions from different sources in the SMAP operating spectrum. However, these source emissions are produced at a small level and required to be well below the source transmitted power. This is still a major concern because these out-of-band (OOB) transmission powers can be at very high levels given the nature characteristic of L-band operators.

Since most out-of-band emissions can be characterized, this makes it easier for conducting analysis on the possible interference that might occur. Transmit and standard properties of these sources were used to gather statistics of RFI level power levels compiled. These simulations provided an understanding to the extent of these power ranges and showed that the range can reach very high levels of power in some situations. In addition, the simulations showed that RFI corruption could occur with frequencies and power levels sufficient to affect SMAP science products. Because radar transmissions are pulsed, a pulse detection method involving oversampling the radiometer measurements in time (as conducted for Aquarius) was originally proposed as sufficient to mitigate these concerns.

2.2 RFI Ground and Space Observations

One way to observe RFI interference activity is through incorporating information from ground-based observations. These observations may be limited in time and space, but are able to provide specific helpful details regarding the frequencies, power levels, bandwidths, and even modulation types of the interference detected. These are valuable in describing the characteristics of the RFI environment that is dealt with. These ground base observations require advanced equipment for accuracy and computation and it's desirable if this equipment is also capable of operating as a radiometer to assess radiometric performance in the presence of interference. The SMAP mission used an advanced radiometer that included systems with digital backends for RFI detection and mitigation [16]. The results of these analyses showed the clear presence of spurious emissions (or perhaps sources operating illegally) in the protected band, and also provided information on the wideband pulsed nature of OOB emissions of L-band radar systems. Sources observed within the protected band from the ground-based locations considered were primarily of the narrowband type, most likely spurious emissions from nearby communications systems occupying bandwidths of 1 MHz or less. Such source types would be difficult to detect and mitigate with a pulse detection strategy alone.

Since ground based observations were found to be limited in time and space, it was necessary to add an extra dimension of observation. Aircraft based observations was an alternative method used to gather more data regarding RFI interference as more space coverage became accessible. Previous datasets from previous similar missions like SMOS were also examined and added to the process of evaluation [16]. Figure 2 illustrates the RFI present in data collected over approximately 24 hours of flight at the L-band. Two digital and one analog backend systems were included for RFI detection and mitigation analysis. Source characteristics that were

observed included pulsed, narrowband (or continuous wave), and wideband types. These results were outputs of both out of band emissions and possible illegal operating sources inside the protected L-band.

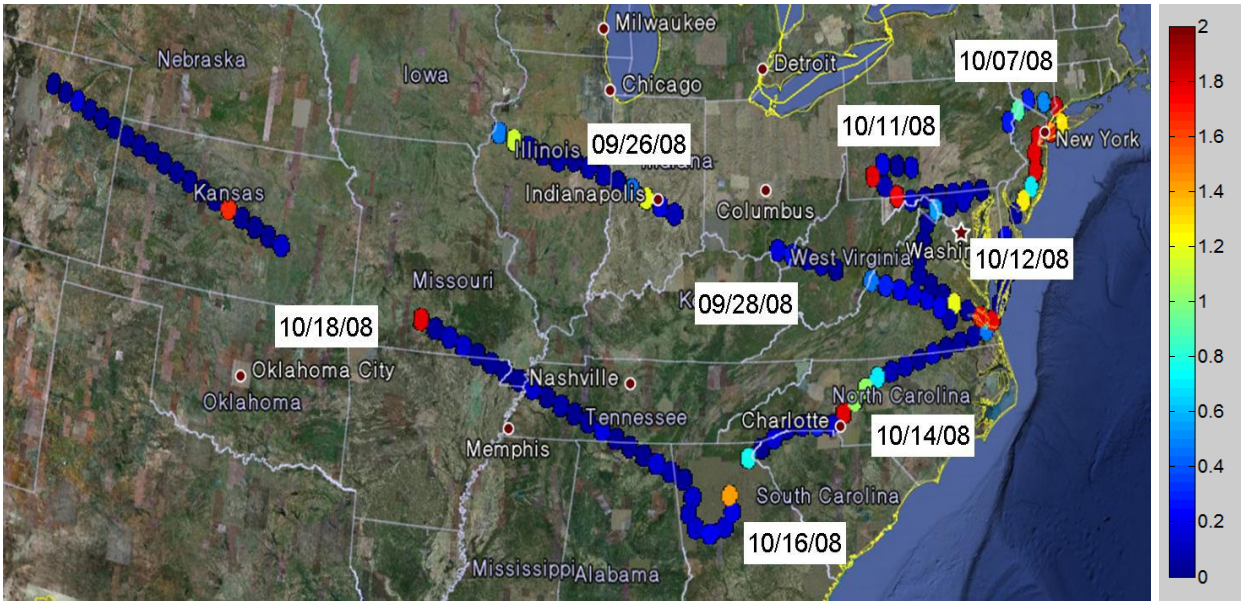


Figure 2: Color Scaled RFI Data Over Flight Path [16]

2.3 RFI Satellite Observations

A further expansion to the observation of the RFI environment can be done through satellite measurements. Satellite observations were able to expand space observation from North America alone, to worldwide global measurements. Previous NASA missions operating in the same bandwidth such as SMOS and Aquarius provided radiometer measurements that allowed global information to be obtained. Both missions showed significant levels of RFI corruption. Results and observations from these global satellite measurements showed that North America encountered a reduced level of RFI corruption in comparison to Europe and Asia.

The SMOS mission products were able to provide details of brightness temperature by itself rather than an estimate of RFI levels. An analysis of temperature brightness of the SMOS mission was thus conducted since detecting low-level RFI in SMOS is difficult. This will narrow the focus down to the presence of large sources that are clearly distinguishable from the norm average measurements. It is also noted that RFI corruption of SMOS observations is caused by RFI sources within radiometer main beams and artifacts associated with imaging aliases or sidelobes [3]. Thus, RFI source occurrence probabilities obtained from SMOS measurements are likely overestimates of the presence of "true", actual sources.

Unlike the SMOS mission, the Aquarius radiometer ground processor includes a pulse detection algorithm (based on its oversampling the radiometer measurements in time), allowing RFI contributions (including low level contributions) to be estimated. While this algorithm is well suited to the detection of pulsed interference as from radar transmissions, it is largely inapplicable to low-to-moderate level continuous wave type sources. Therefore, RFI information obtained from Aquarius should be expected in some aspects to be a lower bound on "true" source properties. As in SMOS, RFI contributions from sources in the sidelobes of the Aquarius antenna are also observed.

There are special considerations to include when comparing previous missions to the current mission of SMAP. Difference in spatial resolutions and temporal sampling should be examined. It turns out that SMOS and SMAP have close coverage and spatial resolution, thus scaling has a negligible effect on final statistics. On the other hand, the Aquarius mission was found to have approximately 100 km spatial resolution which results in a footprint area about five times larger than SMAP. Thus scaling RFI power levels by increasing it by a factor of 5 would achieve a more realistic estimation for SMAP mission purposes. Note that this is a simpler explanation and

logic behind RFI examination, because the process of scaling requires a higher level of detail in computation and will not be discussed in this thesis.

2.4 Approach & Collective Characterization

RFI detected from the SMOS mission detected that most interference encountered occurred at temperature brightness below 300 K. SMAP has conducted it's own measurements and tests to ensure consistency with data presented. SMAP has shown similar behavior in having a close to logarithmic relationship of percentage of RFI detected versus temperature brightness. A clear distinction between low and high level of RFI levels was established, however, low level RFI wasn't predictable due to the lack of information for it. There are many clear uncertainties remaining in producing a robust prediction of the SMAP RFI environment, as well as potential evolution of the RFI environment in time, the SMAP mission is currently adopting a conservative approach by utilizing the SMOS extrapolation to characterize RFI statistics over the SMAP soil moisture retrieval area. This approach, for example, specifies that 10% of SMAP observations will contain RFI of 20 K or more. The corresponding RFI amplitude probability density function contains a significant presence of low-level RFI of amplitude ~ 1K or less, and represents a challenging environment for RFI performance tests. Simulations of SMAP radiometer RFI detection and mitigation performance in this environment will be discussed further in Chapter 5; current results of such simulations show success in meeting science requirements for this environment (given a wide range of RFI source types) and indicate a high likelihood of meeting science requirements once on orbit.

3. SMAP HARDWARE COMPONENTS

A big picture display of the SMAP satellite observatory components are shown in figure 3.

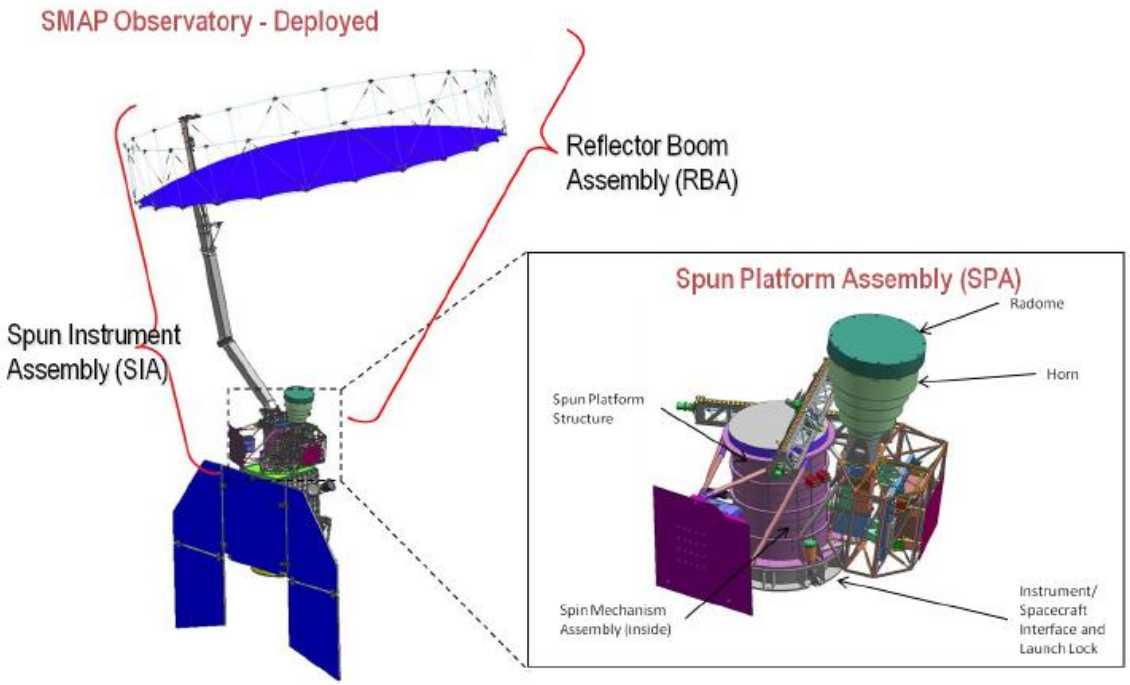


Figure 3: SMAP Instrument Components [2]

The figure contains a reflecting antenna that rotates around the speed of 14.6 rpm. The integration time and footprint size of the radiometer are based on that speed. It is not possible for a single satellite to view the entire globe simultaneously; the "swath" of a satellite is used to describe how much of the Earth's surface is covered in one "measurement". The swath of the SMAP radiometer is around 1000 km as shown in figure 4 on the next page. The antenna footprint will represent the amount of space covered through a given period of integration time. The blue circle in figure 4 illustrates the circular rotation across the swath of the satellite at the speed mentioned of 14.6 rpm.

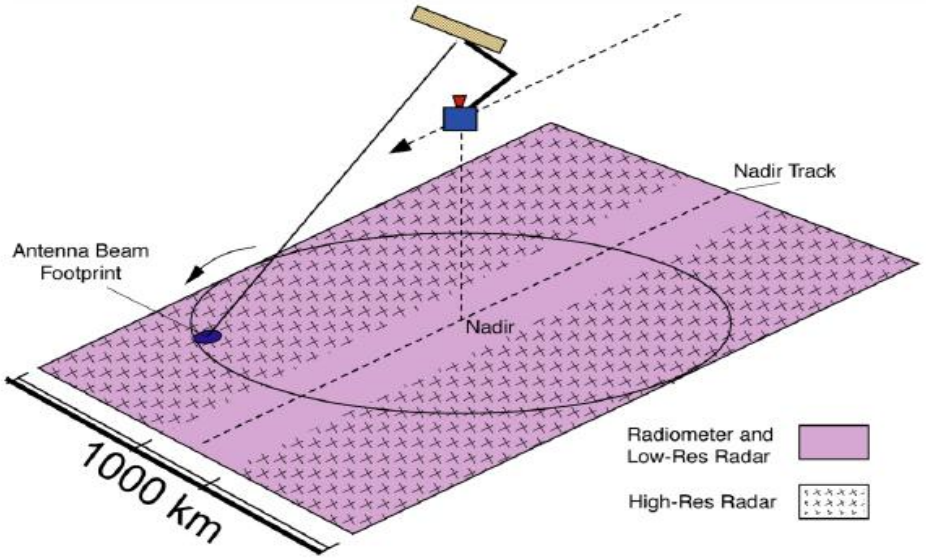


Figure 4: SMAP Radiometer Swath Showing High & Low Resolution Radars [2]

The radiometer operates at a bandwidth frequency of 24 MHz centered at 1.4135 GHz. A significant aspect of the SMAP radiometer design is the utilization of an advanced digital signal processing (DSP) backend. Most radiometer detectors use what is known as square-law diode detectors and analog integrators. Instead of using these integration circuit parts, the radiometer digital electronics (RDE) samples the receiver signals and performs the detection using DSP algorithms. The digital system also allows the data to be integrated synchronously with the SMAP radar pulse-repetition interval (PRI) of approximately 350 μs. The RDE outputs data sufficient for the ground science algorithm to compute the full stokes vector and kurtosis for vertical and horizontal polarizations within the 24-MHz fullband (time-domain) channel and 16 subdivided 1.5-MHz subband (frequency) channels.

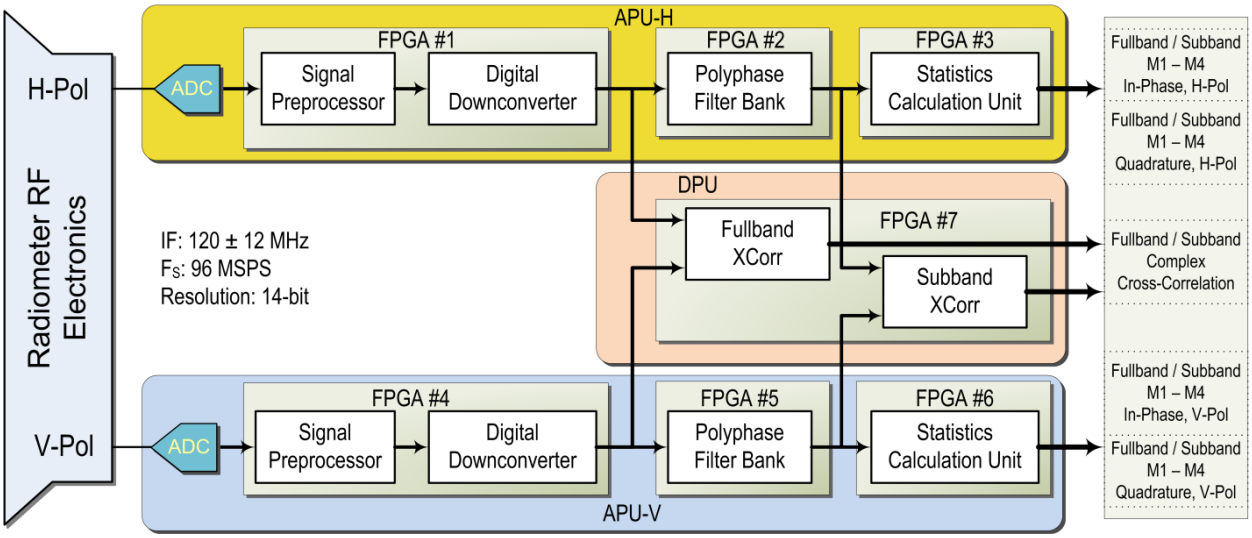


Figure 5: SMAP DSP block diagram showing the 7 FPGAs that perform signal processing and form output data products[1]

The system consists of two analog processing units (APU), horizontal and vertical, and one digital processing unit (DPU) (see Figure 5). The APU cards are mixed-signal cards, containing analog-to-digital converters and synchronously digitizing the pre-detected radiometer signals at 96 MSPS and 14 bits of resolution. The APU cards are interfaced to the DPU via a high-speed digital backplane card. The DPU performs additional signal processing functions and coordinates all radiometer control and command functions, data processing and data packing.

The APU and DPU form what is known as a field programmable gate array (FPGA) based digital signal processor (see Fig. 5). Radiation-tolerant FPGAs are used to make the system robust to failures potentially caused by in-orbit radiation effects. This will help with unpredicted environmental variations during the flight process. On each APU, the radiometer signal is digitized and has its DC offset removed by the signal preprocessor module. Next, the signal is processed through three major modules consisting of a digital down-converter, “polyphase” DFT filterbank and statistics calculation unit. Each of these three modules resides in its own FPGA.

First, down-converted signals products are fullband signals. Second, fullband signals that have been processed by the filterbank are called subband signals. The APU cards compute the first four raw sample moments of the fullband and subband signals. The first two moments are used to estimate antenna brightness temperature, while the first through fourth sample moments are used to estimate the kurtosis for each integration period. The output signals of the digital downconverter and polyphase DFT filterbank for each APU are also routed over the backplane to the DPU, where their complex correlation coefficient is computed to form fullband and subband data products used for the third and fourth Stokes parameter measurements. These modules and processes will be discussed briefly.

1) Digital Downconverter (DDC): The radiometer has a digital bandwidth of 24 MHz by the time it reaches the RDE. To conserve DC power, the bandwidth is passed through a baseband that will ultimately lower the processing. It performs the downconversion on each polarization channel as well. The output of the DDC output will then be a fullband output signal that will travel to a "filterbank" and a "complex correlator".

2) Polyphase Filter Bank (PFB): The filterbank surves the purpose in localizing RFI located in the flight path using a joint time-frequency algorithm. They are able to periodically time vary coefficients so that all 16 subband channels can share the same processing engine and reducing the FPGA logic. The frequency response of the 16 channels can be shown in Figure 6 [1]. Each channel was designed to overlap with its two adjacent channels, as is normally found in spectrum-analyzer applications of filterbanks. Each channel crossover point occurs approximately -2.745 dB. Variations in frequency reponse during flight path can be automatically adjusted.

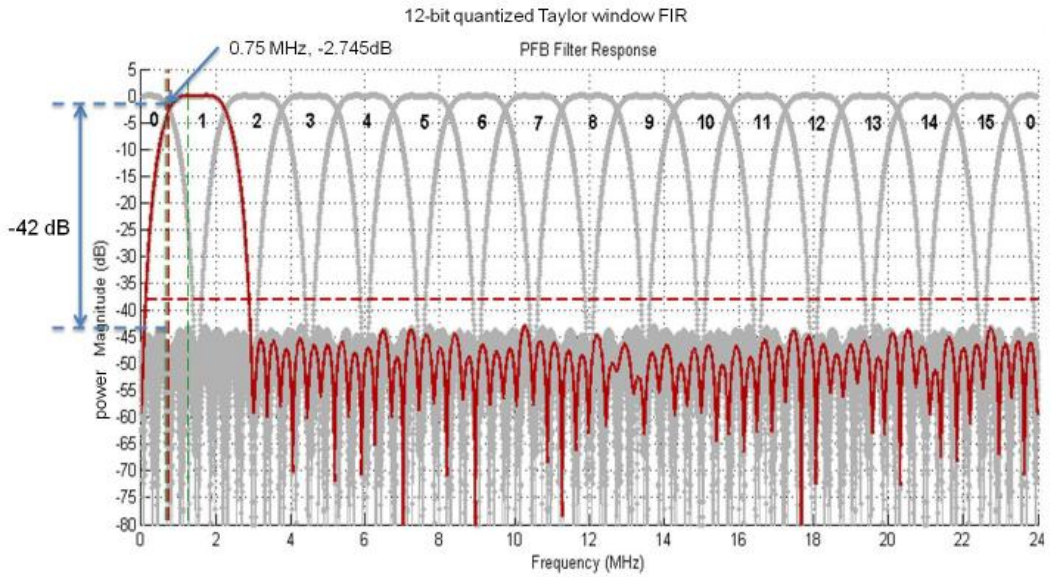


Figure 6: SMAP RDE Frequency Response [1]

3) Complex Cross Correlator (XCORR): The radiometer also computes the complex cross correlation coefficient between the vertical and horizontal polarizations for measuring the third and fourth Stokes parameters. Both fullband and subband correlations are performed on the DPU. Subband cross-correlations are computed for each of the 16 channels using a shared hardware complex multiplier. All subband signals are correlated separate from the fullband signals, using a highly multiplexed complex multiplier and data buffers to organize the samples from the interleaved subband signals.

4) Statistics Calculation Unit (SCU) and Science Telemetry Output: This is the instrument that helps in computing the first to four moments discussed in Chapter 1.5. For each sample signal, the in-phase and quadrature components included, the moments are processed. Sample moments will also be computed in the hardware to output Kurtosis which will become useful for future analysis. These moments aren't centralized however, and thus attention must be given when retrieving those products for analysis.

The outputs of the SCU and XCORR are packetized and communicated to the ground during periodic downlinks. The RDE is set to over sample the angular Nyquist rate of the radiometer by 44 and 11 times for the fullband time domain and subband frequency domain, respectively. The "time-domain" detection will be directed to the fullband antenna temperature products of the radiometer. Fullband products are evaluated in time domain and are integrated at approximately ~350 μs. Measurements from fullband products will include both horizontal (H) and vertical (V) polarizations. Each footprint time sample contains 44 pulse repetition intervals (PRIs). The passband contains 16 subbands that are integrated over a period of ~1.4 ms packets. Each footprint contains 11 packets, with each packet containing 4 PRIs. Horizontal and vertical polarizations are also included in antenna temperature data collection. Further details and reference can be shown in Figure 7 and Table 1.

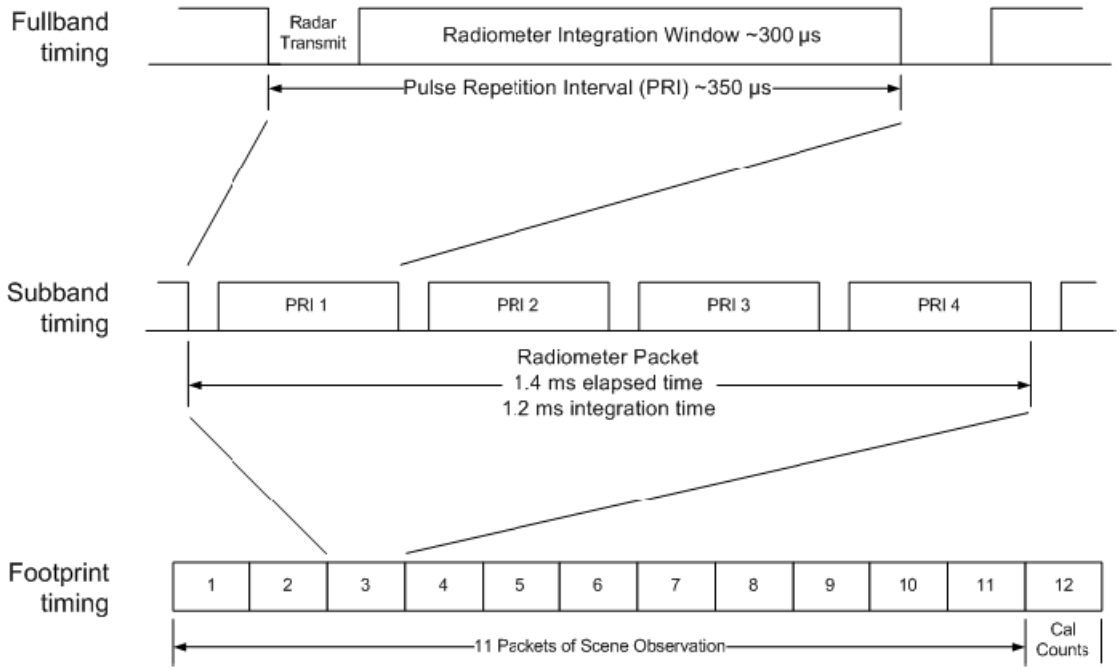


Figure 7: Radiometer Timing and Sampling Sequence [1]

Polarization or Stokes parameter	No. of freq. channels	Samples per footprint	Baseband I/Q components	Raw moments or cross correlations
		Fullband Channels		
V	1	44	2	4
H	1	44	2	4
3	1	44	1	1
4	1	44	1	1
		Subtotal of measurements:		792
		Subband Channels		
V	16	11	2	4
H	16	11	2	4
3	16	11	1	1
4	16	11	1	1
		Subtotal of measurements:		3168
		Total measurements for single footprint:		**3960**

Table 1: Single Footprint Measurements for Fullband & Subband Channels across all Polarizations [1]

The following section will discuss how these output products are processed to produce two main level products to use to evaluate radiometer digital electronics and overall system performance.

4. RADIOMETER SOFTWARE PRODUCTS

Collected the radiometer products for analysis and observation is an essential process to a successful mitigation system. The products can be manipulated to evaluate the radiometer digital electronics (RDE) as well as the overall system performance (mitigation and processing effects). The products retrieved from the radiometer can be thought of as “raw” or Level 0 data. The conversion process is a two staged procedure. The first stage is converting raw data taken from the RDE into what is named ‘Level 1A’ data through MATLAB coding. The second stage is

processing ‘Level 1A’ to ‘Level 1B’ products also using MATLAB coding. Both these level products are useful for different applications and tests. A more detailed discussion will be included in the following sections. Figure 8 illustrates a block diagram that shows the process with detailed outputs to ease the understanding of the route taken. The ancillary data are prior information gather about the systems behavior and thresholds set. These are used for calibration when transforming Level 1A data into Level 1B.

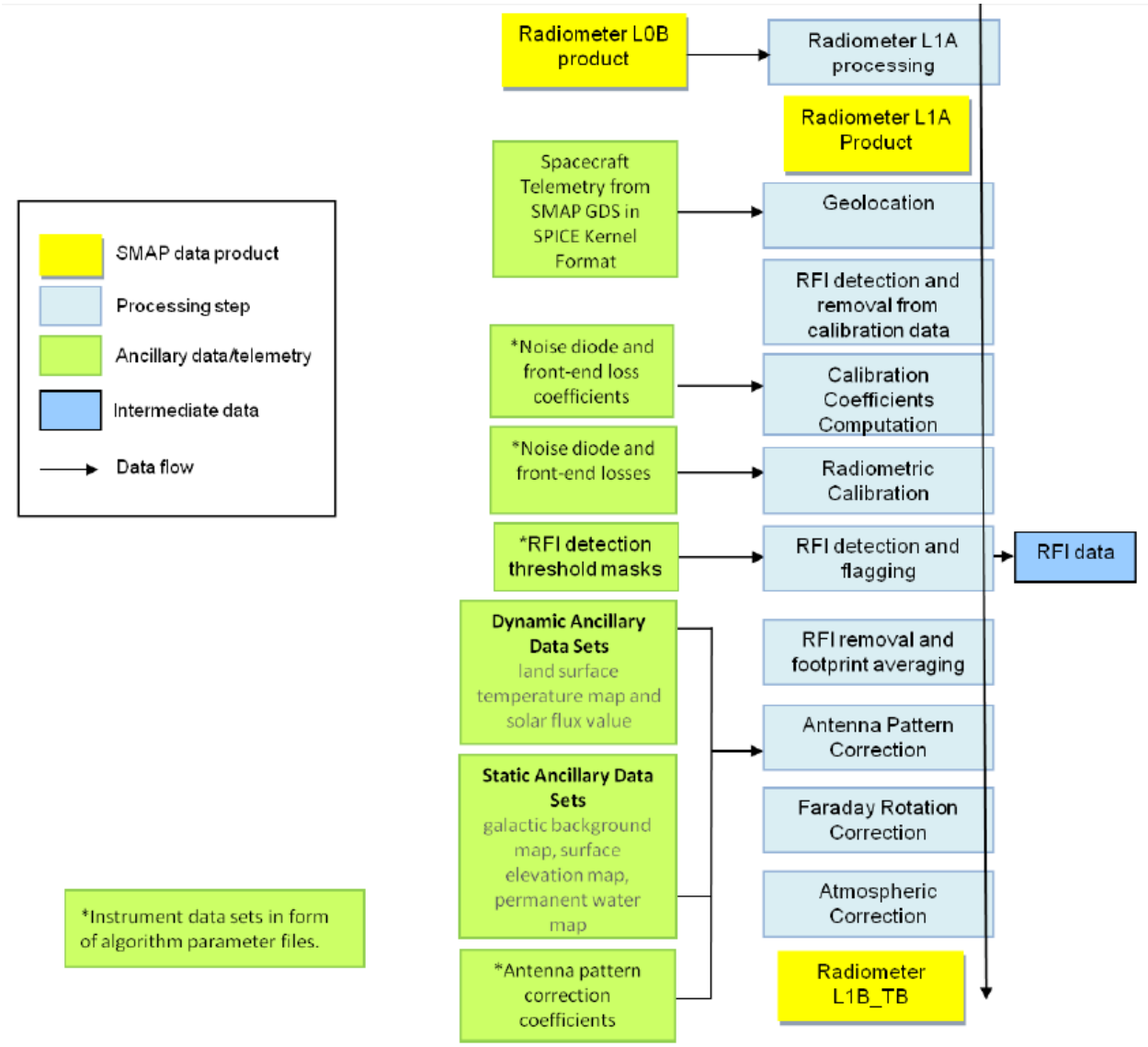


Figure 8: Block Diagram of Software Level Output

4.1 Level 1A Products

Level 1A outputs contain products that are symmetrical to those outputted by the radiometer. Thus, implementing a system that mimics the behavior of the radiometer outputs is essential. We know from the radiometer hardware components discussed in the previous chapter, that the radiometer outputs fullband and subband component products of first to fourth moments, as well as the cross correlation producing the third and fourth stokes.

As shown from the block diagram in Figure 9 below, Level 1A products output data interpreted in the time-domain. These products parallel to the radiometer are labeled "FullMomGroup. All antenna data contain the first through fourth moment's labeled "m1-m4" of horizontal and vertical in phase and quadrature (HI, HQ, VI, VQ). As well as the variance of third and fourth stokes antenna measurements labeled "t3-t4". There are also other measurements included such as noise diode, and cross noise diode measurements of the antenna radiometer, these are labeled antenna (nd, xnd).

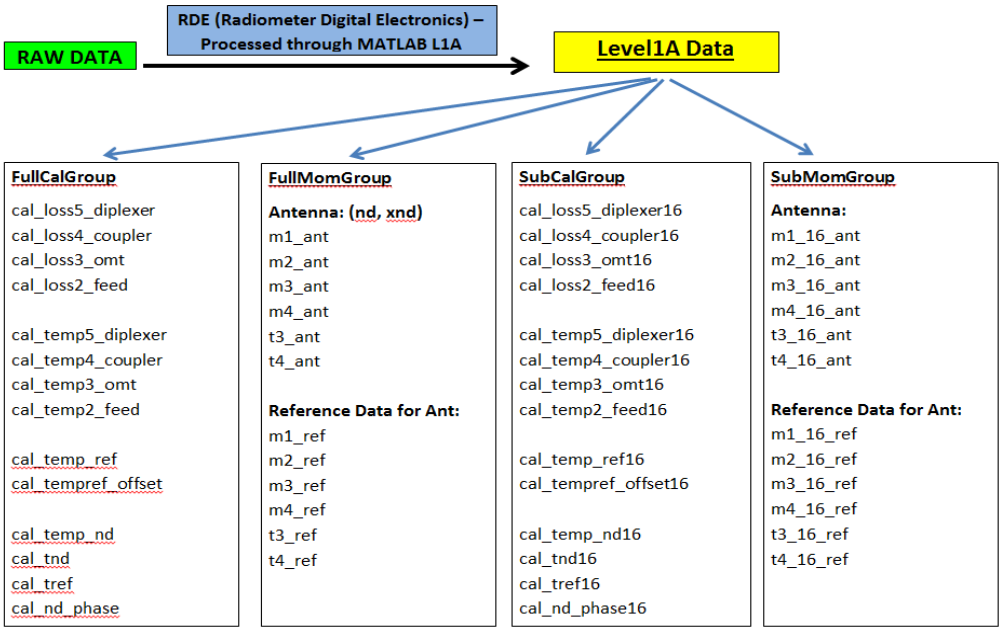


Figure 9 : Processing Raw Data to Level 1A Data with Products

Associated with the heart of radiometer products, ancillary information is added to these products to ease the process of calibration converting Level 1A to Level 1B. These products can be shown in Figure 9 under the label “FullCalGroup”, ‘cal’ referring to calibration. The details of these products are not necessary for our purposes, however, it is useful to know that there purposes is to serve the following step of calibration outputting the final brightness temperatures. Some of the data needed for calibration includes reference load temperature, noise diode and cross noise diode effective temperature, front end losses, and correlation phase correction. The Antenna contains approximately ~10560 measurements and around ~480 reference load measurements.

4.2 Evaluation of Radiometer Digital Electronics using L1A

Level 1A data products are reconstructed, unprocessed instrument data at full resolution, time-referenced and annotated with ancillary information. The Level 1A product is a time-ordered series of instrument counts and includes housekeeping telemetry converted to engineering units for each scan. Generally, previous processing code has been put in place prior to my start work on the project. However, slight variations and adjustments to the processing code were necessary for better optimal performance. These slight variations conducted on the processing code were successful in producing Level 1A products. From there, further work on our end can be done to evaluate the radiometer digital electronics (RDE) performance. This can be done through studying properties of variance, skewness, and kurtosis for RFI free and RFI containing datasets. Mean and standard deviation of kurtosis can be compared with theory for pulsed-sinusoidal RFI. Referring back to the properties of skewness and kurtosis mentioned in Chapter 1, it becomes easier to understand that the kurtosis and standard deviation of kurtosis should be relatively

constant over time. This will result in more accurate measurements relying on a more uniform Gaussian distribution.

Previous evaluation to kurtosis on the SMAP radiometer showed variances in kurtosis across time as can be seen in Figure 10. This test was done over the "engineering test unit" ETU model for preliminary investigation and analysis. This was particularly helpful in understanding that the

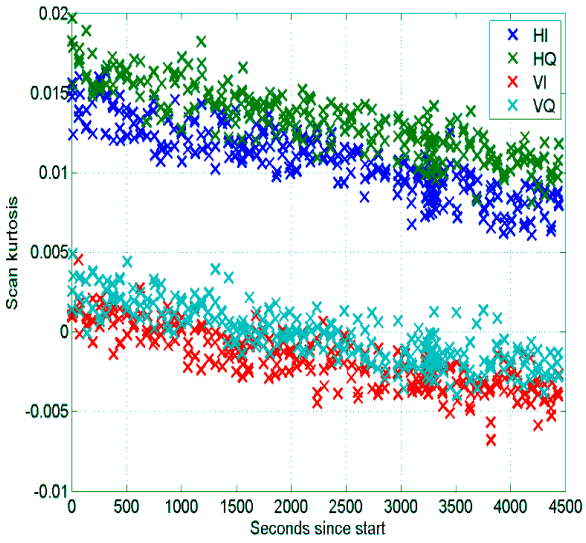


Figure 10: Kurtosis Variation (ETU Data)

radiometer isn't working to full efficiency. A reported back problem was held to investigate the causes of such variances, and it was determined that adjusting "pointer" issues into the physics of the radiometer helped achieved a more consistent variation over time.

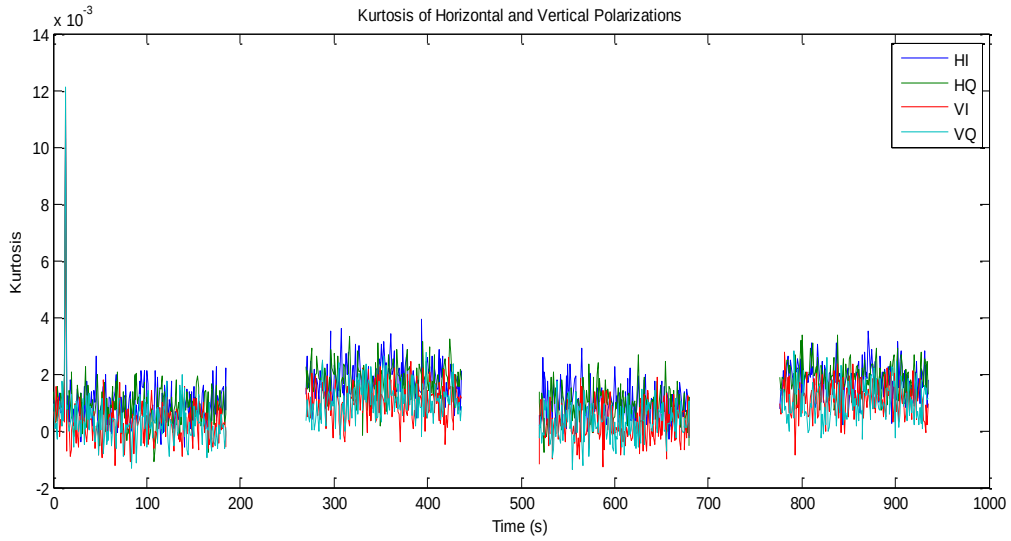
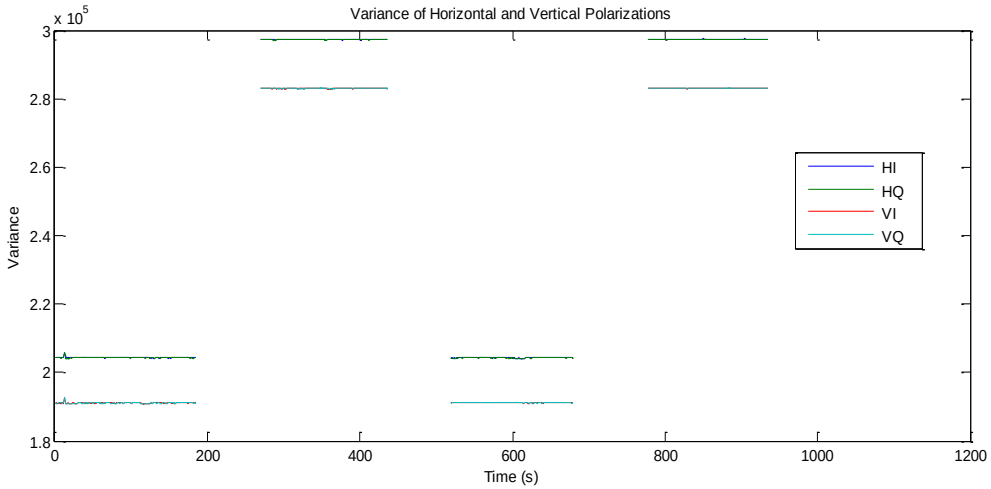


Figure 11: Kurtosis Variation (Flight Data)

Later evaluations were conducted during a second stage testing labeled "flight data" testing. During that testing, properties of kurtosis led more consistent and promising trends over time as shown in Figure 11 on the previous page. Note that the data has is discontinuous for periods of time where this corresponds to switching between RFI ON and OFF. Variations in kurtosis seem to be consistent regardless of presence of RFI. Variance and skewness followed the same results of having variations at a minimum.



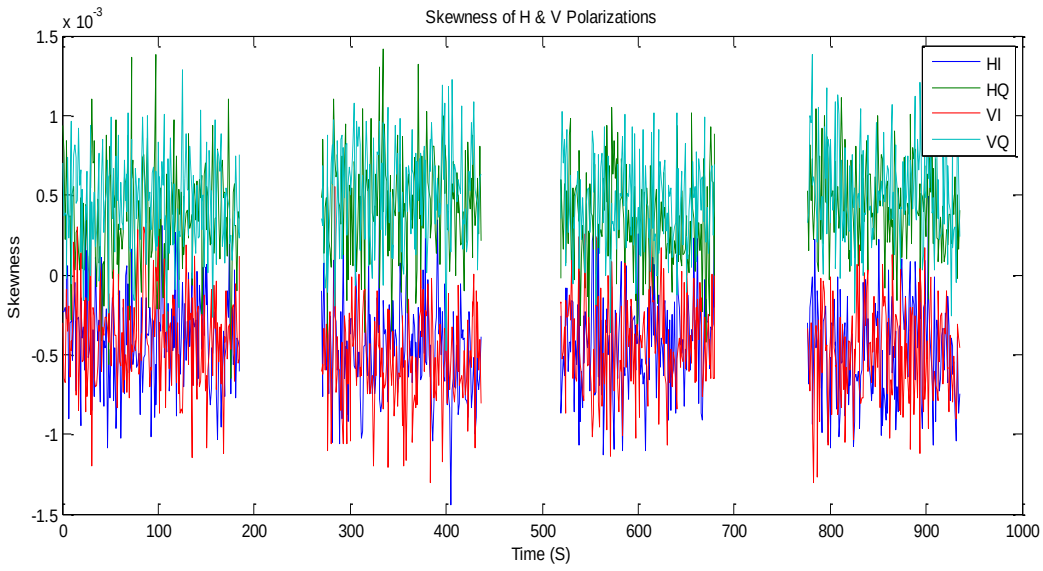


Figure 12: Variance and Skewness Properties of Horizontal and Vertical Polarizations

4.3 Level 1B Products

After processing Level 1A data products, calibration can be done through the ancillary data provided. The final products will include measurements in units of temperature brightness. Outputs correspond to Level 1A files and are a footprint scale or, 11 ms observation as discussed in Chapter 3. Temperature brightness is computed before and after RFI mitigation to show the effects of the RFI algorithms implemented into the system. Mitigated data is obtained through integrating temperature brightness over the spectrogram which contains 16 frequency channels and 11 packet resolutions. The products can be shown in Figure 13 where there are two separate set of data. The first, labeled under "TAs" represents measurements conduction with RFI included. Temperatures are conducted in both horizontal and vertical polarizations as well as in time and frequency domains.

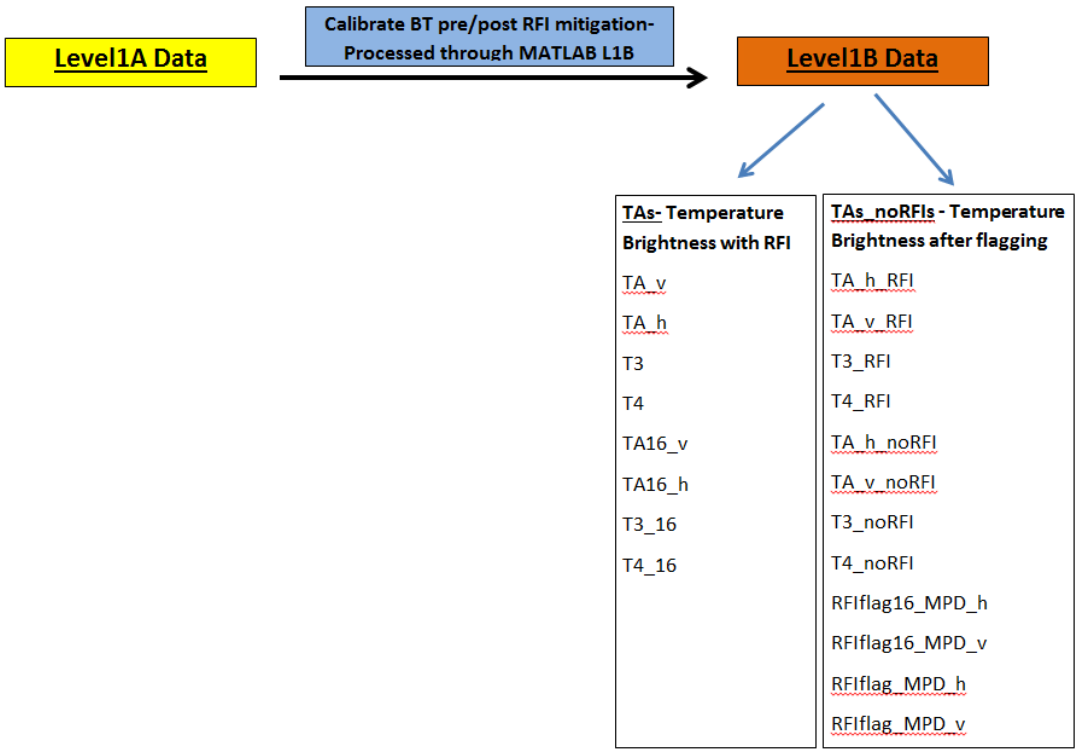


Figure 13: Calibrating Level 1A Data into Level 1B

4.4 Evaluation of Overall System Performance using L1B

Similarly to evaluating Level 1A products, variations in calibration code and systems were necessary to achieve desired outputs. The analysis performed on Level 1B products can be used to understand how well the system responds to RFI injected sequences. The SMAP system injects a sequence of RFI ON followed by a sequence of RFI OFF. Typically, both these sequences last around 30 seconds. These intervals can be analyzed to determine whether successful mitigation was performed on the system.The figure below shows the two case scenarios of RFI containing and RFI free data sets. As seen, the performance of the RFI algorithms as well as the radiometer DSP algorithms performed well. Under both cases, the mean of temperature brightness seemed to output a fairly close average measurement. 263 GB of "Level 1B" data is used to evaluate the overall system performance under different conditions

of RFI.

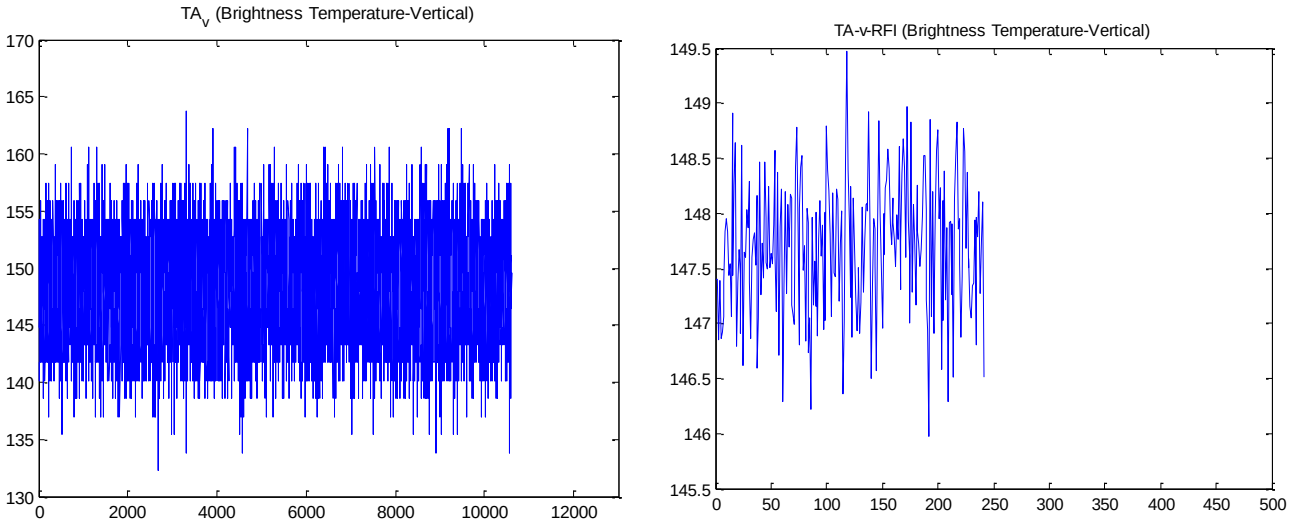


Figure 14: Vertical Brightness Temperature with no RFI (left) and with RFI (right)

5. RFI DETECTION ALGORITHMS

RFI can corrupt measurements through either non-physical interference leading to negative valued injections, high powered (positive valued) interference, or sporadic dry biases. The sporadic result can be the most harmful because undetected RFI could be mistaken for environmental signatures. Thus, there exists a compelling justification for aggressive RFI mitigation in SMAP.

5.1 Pulse or Time-Domain Detection

The pulse or time domain detection algorithm searches in the time domain for increased levels of observed antenna temperatures above that produced by geophysical properties. No periodic properties of the RFI are assumed for this type of detection, which is why it's sometimes referred to as asynchronous pulse banking. Time domain detection algorithms are generally directed towards RFI with large amplitudes and short duration and duty cycles, or "spikes". The L-band

has been recorded to observe pulsed type RFI's are ranges from 2 to 400 μs in length at time separation of 1-75 ms. Detection of these pulses require an evaluation of the statistics and probability theory of the actual data retrieved. From the reference in the introduction, one can find that the properties of mean and standard deviation can become useful to apply RFI algorithms. Usually mean and standard deviation are computed by eliminating the largest N% of samples to prevent high false alarm rates. A threshold is put so that if the "time-domain" sample is a certain number of standard deviations above the mean, the algorithm "flags" it as RFI. The standard deviation is determined through evaluation the system temperature prior to testing to reduce the FAR. When a "time-domain" sample is detected and flagged, previous and next samples are also flagged to ensure neighboring RFI samples aren't affecting measurements.

The "time-domain" detection will be directed to the fullband antenna temperature products of the radiometer. Fullband products are evaluated in time domain and are integrated at approximately ~350 μs. Measurements from fullband products will include both horizontal (H) and vertical (V) polarizations. Each footprint time sample contains 44 pulse repetition intervals (PRIs). Time-domain detection will be evaluated over one footprint and expressed as,

$$\boldsymbol{T_A(t) - m(t) \geq \beta_{td}\sigma_{td}(t)} \quad \textbf{(5.1)}$$

where *m* is the mean of samples; σ_{td} is the standard deviation of measured samples; β_{td} is the threshold of data; and T_A is the measured temperature sample. When calculating the mean, N% number of samples are removed as previously discussed, and a choice of N= 10 has been selected to avoid biasing. This window contains samples from the footprint under observation as well as samples before and after. The number of samples can either be 1, or may exceed to 2 and 4 for prolonged integration purposes. An example simulation of time domain RFI detection and

mitigation is shown in Figure 15.

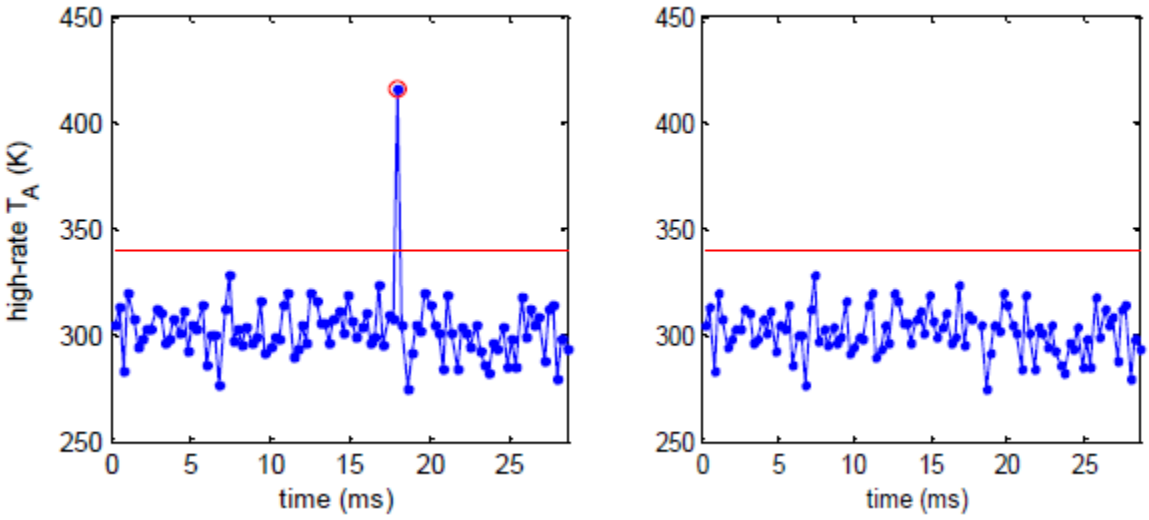


Figure 15: Time-Domain Detection Illustration [1]

As you can see from the figure, the red horizontal line represents the threshold value selected. When a "spike", or high amplitude low pulse width RFI, occurs, the time-domain baseline detection is able to flag and negate the interference by mitigation the collected sample, with previous and next samples averaged out. The final plot on the right will represent a more accurate picture of the actual data measured.

5.2 Cross- Frequency Detection

The cross-frequency detection algorithm can be thought of as a parallel detection method to the pulsed method. The difference between the two is that cross-frequency detection algorithms evaluate samples in the frequency domain rather than the time domain. The SMAP radiometer contains $2^4 = 16$ frequency channels. The cross-frequency detection seems to perform best with narrow band RFI sources that have close frequencies to the samples measured. However, there isn't a specific direction or RFI source that cross-frequency is focused on. A similar computation

of mean is calculation by eliminating the largest N channels to avoid biasing. The flagging procedure, like time-domain, determines which samples are number of standard deviations above the threshold, determining the FAR. Although cross-frequency isn't necessarily directed towards a specific type of RFI, it seems to be more sensitive to continuous wave RFI relative to other detectors.

The cross-frequency algorithm operates on the subband antenna temperature measurements. The passband contains 16 subbands that are integrated over a period of ~1.4 ms packets. Each footprint contains 11 packets, with each packet containing 4 PRIs. Horizontal and vertical polarizations are also included in antenna temperature data collection. For each time sub-sample the RFI detection occurs similar to time-domain with different threshold, mean, and standard deviation values used. The equation can be expressed as,

$$\boldsymbol{T_A(f) - m(f) \geq \beta_{cf}\sigma_{cf}(f)} \qquad \textbf{(5.2)}$$

Again with the same parameters defined as before, but in frequency domain. The equality resembles the same relationship between the difference between a sample temperature measurement versus mean, and evaluation the value for a given constant of threshold multiplied by the set of samples standard deviation. Any channel which contains TA that are βcf standard deviations above the mean is flagged as corrupted and then removed. Subbands adjacent to those flagged as containing RFI are also flagged as corrupted and removed. As in the pulse detection algorithm, an adaptive standard deviation is not necessary and can be determined in pre-launch calibration. The N number of channels chosen to be eliminated in the cross-frequency case is 2. Thus evaluation will only be conducted through 14 channels to reduce the FAR as much as

possible. After detection and removal of RFI samples, the remaining samples are averaged to replace the removed footprint at the corresponding frequency channels.

The process of detection and removal of a frequency channel can be illustrated in Figure 16 on the next page. The figure represents an example of injected RFI CW frequency at 1413.5 MHz and the mitigation process on the right. As seen, the red frequency subband has been flagged and removed as well as the corresponding neighboring subband channels. The next step process would be to replace the mitigated subband channels with the average samples of the remaining frequency spectrum.

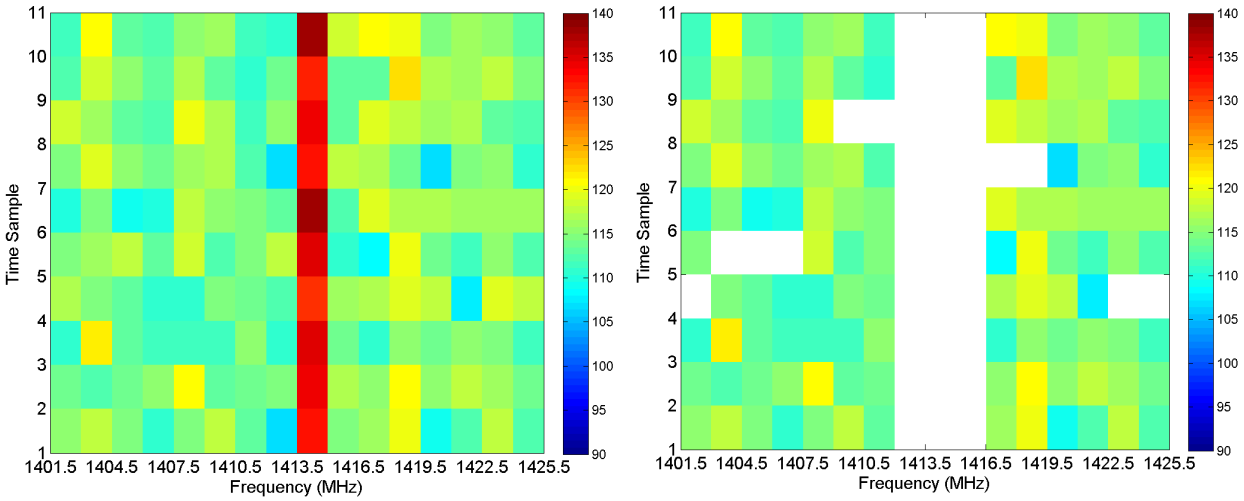


Figure 16 : Cross-frequency Detection Illustration

5.3 Kurtosis Detection

The previous detectors resembled efficieny with definite RFI sources such as high amplitude low bandwidth pulses as well as high amplitude continuous wave pulses. The physical natures of measurements also include random sources of RFI. These sources tend to have unknown characteristics, but merely a source of randomness. Thus the kurtosis detection algorithm drives

us back to the properties of randomness and statistical theory to detect RFI. Thermally generated radiometric sources have an amplitude probability distribution function that is Gaussian in nature, whereas man-made RFI sources tend to have a non-Gaussian distribution [3]. The kurtosis algorithm measures the deviation from normality of the incoming radiometric source to detect the presence of interfering sources.

The kurtosis detection algorithm measures higher order central moments of the incoming signal than the 2nd central moment measured by a square-law detector in a total power radiometer. The Kurtosis referring back to Chapter 1.5 is denoted as the fourth standardized moment. The fourth standardized moment can be written as,

$$\boldsymbol{K = \frac{\mu_4}{\gamma_4} = \frac{E\left[\left(x - E(x)\right)^4\right]}{\left(E\left[\left(x - E(x)\right)^2\right]\right)^2}} \quad \textbf{(5.3)}$$

This denotes the fourth moment of x, divided by the square of the second moment. For algorithm purposes used in the software, kurtosis can be expanded and written in terms of individual moments of x as,

$$\boldsymbol{K = \frac{m_4 - 4m_1m_3 + 6m_1^2m_2 - 3m_1^4}{\left(m_2 - m_1^2\right)^2}} \quad \textbf{(5.4)}$$

The scaled factor subtraction of 3 mentioned in chapter 1.5 can be explained alternatively. The kurtosis equals three when the incoming signal is purely Gaussian distributed and in most cases deviates from three if there is a non-normal (typically man-made) interfering source present. The kurtosis statistic is independent of the 2nd central moment of the signal, i.e., the kurtosis value is not affected by natural variations in the antenna temperature of the scene being observed.

The kurtosis estimate itself behaves like a random variable since it is generally calculated from a finite sample set [3]. Similar to previous detection algorithms, estimates of the kurtosis have a standard deviation associated with them, and there is a corresponding kurtosis threshold for detecting RFI. If the sample size is sufficiently large, which is the case for SMAP measurements, the kurtosis estimate exhibits a normal distribution. The kurtosis detection algorithm will detect the presence of RFI with the SMAP radiometer using the kurtosis statistic. The kurtosis statistic is computed from the first four raw moments of the radiometer signal.

The Kurtosis detection identifies significant statistical variations from the normal purely thermal emission origins that are Gaussian distributed. A threshold of the algorithm still exists and is a parameter than can be adjusted to establish a low false alarm rate, high probability of detection. Unlike time-domain and cross-frequency where each algorithm was limited to one domain, kurtosis detection is set to operate on both fullband and subband data products from the radiometer. The process of mitigation and removal of flagged samples are parallel to those of time-domain and cross-frequency respectively.

The first four raw moments are computed from the products of the radiometer using the logic displayed in Chapter 1. These four raw moments are presented as the inputs to equation (5.4) where m_n represents the n^{th} raw moment. The equation can be used separately on both the in-phase (I) and quadrature (Q) components of the signal for each frequency subband and time fullband as well as each radiometer channel (V and H).

The kurtosis threshold is the deviation from the nominal kurtosis value beyond which a sample is considered to be corrupted by RFI. The value of this threshold determines the false alarm rate and probability of detection of RFI. For each time and frequency subsample, the value of the

kurtosis is compared to the nominal Gaussian value and detection occurs if the deviation from the nominal value exceeds the threshold. Detection using kurtosis occurs if,

$$|K - K_{nom}| > \beta_k \sigma_k \quad \textbf{(5.5)}$$

where , K_{nom}, β_k, σ_k are the measured kurtosis, nominal kurtosis value, threshold value, and the standard deviation of the measured kurtosis respectively. The threshold value usually depends on a predetermined latitude and longitude look up table and will be set prior to launch. Values for the standard deviation will be determined through experiments by the NASA laboratory prior to launch and may be adjusted according to flying conditions eventually. Through experimentation and observation, it turns out that pulsed RFI with duty cycles less than 50% produce kurtosis values greater than 3. While continuous wave (CW) signals as well as pulses with duty cycles greater than 50% suppress the kurtosis below 3. An example from a previous SMAP campaign using kurtosis detection is illustrated in Figure 17. This figure includes both types of injected RFI, pulsed narrow band and continuous wave. As deviations become higher towards the end the spectrum range, kurtosis detection was able to negate and identify these random RFI source signals and produce a standardized or "normal" output.

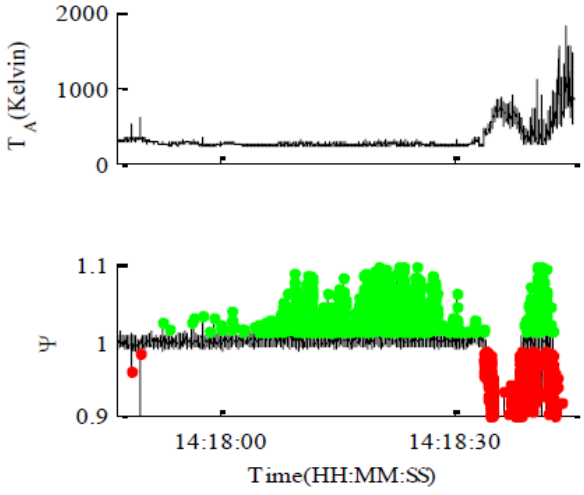


Figure 17: Kurtosis Detection Illustration [16]

The green colors represented “good” data that is passed, while the red colors represent the “bad” data that are flagged. The resulting output would follow a pattern close to the normalized sequence.

5.4 Polarization or “T_3 & T_4” Detection

The final algorithm for RFI detection is included to ensure all measures are taken into account. As discussed in chapter 1.5 the third and fourth stokes parameters are computed by performing the complex correlation of the horizontal and vertical polarizations. Usually, since horizontal and vertical polarizations are highly close in measurements, the cross correlation are nearly always zero. Unless for special circumstances if that isn’t the case, then it can be concluded that RFI is present. Similar to Kurtosis, since third and fourth stokes are statistical measures, analyzed and performance can be performed over time and frequency domain as well. RFI can be detected by looking for unusually large variations in the 3rd and 4th Stokes parameters for the fullband and each of the 16 subbands.

The detection algorithm which uses the 3rd and 4th Stokes parameters is a simple thresholding algorithm which searches for variations greater than a fixed number of standard deviations. The general equality used for polarization is detection is defined as,

$$\left|\boldsymbol{T_{3,4} - T_{nom}}\right| \geq \boldsymbol{\beta_{3,4}\sigma_{3,4}} \quad \textbf{(5.6)}$$

In practice, T_{nom} value is expected to be almost zero over land, small over the ocean and can lead to a further approximation represented by the equality,

$$\left|\boldsymbol{T_{3,4}}\right| \geq \boldsymbol{\beta_{3,4}\sigma_{3,4}} \quad \textbf{(5.7)}$$

The output of each detection algorithm will be a detection flag = 1 if RFI is detected in the sample under test and a detection flag = 0 if RFI is not detected in the sample under test. These RFI flags will then be the inputs to the RFI mitigation or removal algorithm to be discussed in the next section. The RFI flags from all the detection algorithms will be stored in a separate database which can later be used to determine prevalence of RFI geographically.

5.5 Maximum Probability of Detection (MPD)

This section serves the purpose of explaining how the logic behind combining the different RFI detector works. We now have four different (pulsed, cross-frequency, kurtosis, and polarization) baseline RFI detector that we wish to combine. Each of the given detectors output a set of flags corresponding to either fullband data or subband data. The final RFI detection algorithm combines these four set of RFI flags to produce a single maximum probability of detection (MPD) algorithm. The MPD is set in a way that minimizes the probability of missing the detection of RFI.

The SMAP radiometer baseline detection algorithms is capable of flagging a range of RFI sources including temporal "glitch" or "spike" detection, frequency domain outliers, random non-Guassian values, and unusually high polarized detections. To maintain consistency with the decision of a "conservative" approach mentioned in chapter 2, the composite MPD RFI detection algorithm is a logical OR of each of the individual RFI detection flags. Since the probability of missed detection is only partially correlated between individual flags, this can result in the detection of RFI by the MPD algorithm that was missed by one or more of the individual algorithms. On the other hand, due to the logical OR operation, no RFI that is detected by any individual algorithm can ever be missed by the MPD algorithm. For this reason, the MPD flag

minimizes the probability of missed detection given the available individual flags. The evaluation of the RFI baseline detection algorithms will be analyzed in the following chapter.

6. RFI DETECTION PERFORMANCE

The SMAP mission team has raised an area of interest regarding the efficiency of the satellites' RFI algorithms and detectors. Extreme efficiency and precision are needed to satisfy accuracy of data information retrieved regardless of RFI existence. The purpose of the study is to investigate the performance of the different detectors with varied RFI inputs. This will be done through generating receiver operating (ROC) curves to determine threshold levels to be used during ground processing and thereby optimizing the algorithm. The receiver operating curves were generated from flight model test data during thermal vacuum. SMAP holds four different baseline detection algorithms that are processed through a maximum probability of detection (MPD) logic explained in the previous chapter.

For all the four different baseline detection, they each contained a common parameter which was the threshold parameter β. The choice of β determines where the time domain detection algorithm will operate on the receiver operating curve (ROC), thus establishing its false alarm rate (FAR) and probability of detection (PD).

6.1 Receiver Operating Characteristic (ROC) Curves

The receiver operating characteristic curve is a graphing plot that reflects the performance of a given system under different varying thresholds. ROC curves reflect how well a given system is operating under different sets of probabilities. Each point on an ROC curve can be represented by the set {FAR, PD}. Where FAR is the "false alarm rate" (or false positive rate) and PD is the "probability of detection (or true positive rate). For our case in analyzing the RFI baseline

detectors, the false alarm rate gives the rate at which that the RFI detector system had classified a *correct* measurement as RFI. Whereas, the probability of detection gives the rate at which the RFI system had classified a *false* (or incorrect) measurement as RFI as illustrated in equation (6.1).

$$PD = \frac{Positives\ Correctly\ Classified}{Total\ Positives}; \quad FAR = \frac{Negatives\ incorrectly\ classified}{Total\ Negatives} \qquad (6.1)$$

Thus, one would naturally desire the RFI detection systems to detect *all and only all* of interference giving a PD of 100% and FAR of 0%. Clearly, that's an ideal case that wouldn't be possible to match. However, measures can be taken to improve system detection percentage values. Each set of points of {FAR, PD} are created given a threshold level. Varying threshold levels will thus create a continuous plot of multiple {FAR, PD} sets where FAR and PD values will range from [0-1]. An example of a set of points on different ROC range can be displayed in the figure below.

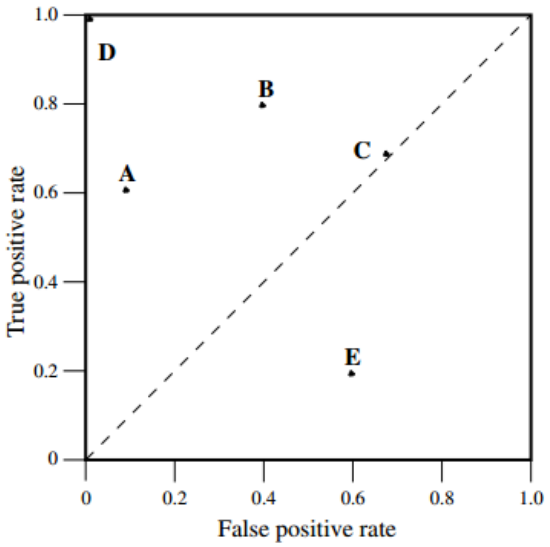


Figure 18 : ROC Range showing 5 point sets of {FAR, PD} [6]

There are different points to note on an ROC graph and will be represented through the RFI detection algorithm system for parallel clarity. The set point (0,0), lower left, represents the point of never issuing a detection system onto the data sets. In other words, this means that the threshold will be set extremely high and thus passing all data through the system without detection of RFI. The opposite strategy of unconditionally issuing positive flags, will set the point (1,1) at upper right of the figure. In terms of RFI detection algorithms, this will mean setting the threshold extremely low so that eventually everything is flagged, where the false alarm rate also reaches a value of 1. Now it can become clear that varying the threshold will produce set point values going from (0,0) to (1,1). Notice the dashed line will represent the worst ROC curve case of always having a 50/50% between detecting and not detecting an RFI source. Whereas the ideal case would represent the [(0, 0) - (1, 0)] line up the y-axis, then the [(1, 0) – (1, 1)] line across the x-axis.

Informally, one point in ROC space is better than another if it is to the northwest (FAR is lower, PD is higher, or both} of the first. Thus the point D would be a better choice than all the rest on the graph. However, it becomes unclear whether it's "better" to operate at point A or B because there's a tradeoff between the probability of detection and false alarm rate. Point B seems to have both higher PD and FAR, where choices of B over A would tend to fall more into the "conservative" approach taken in SMAP. The MPD detection algorithm has indicated that it would like to maximize the probability of detection as much as possible, even if a tradeoff is a higher FAR. In the end, the choice of operation returns to the designers' application and specific situation applied. Theoretical RFI detectors ROC curves were generated and are shown below, but aren't based on actual data collected.

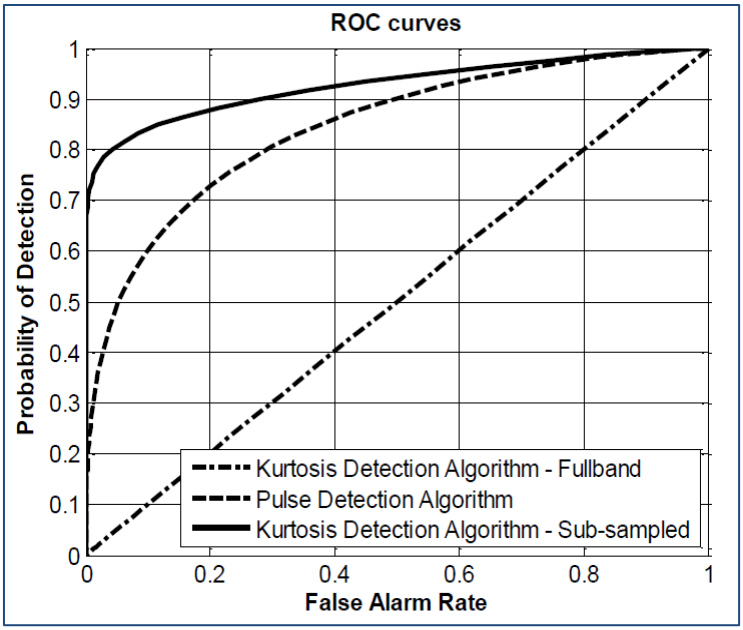


Figure 19: Theoretic ROC Curves [3]

6.2 Area Under an ROC Curve (AUC)

An ROC curve represents a systems performance through a fixed range window. Given that all ROC curves will travel from the point (0,0) to (1,1) or vice-versa, ensuring a continuous line between the two is a definite. A common way to classify different system performers, or in our case, RFI detection performance, is to reduce down the ROC performance into a single value representing expected performance. This could be achieved through computing the area under the ROC curve, or commonly known as the AUC. Since the ROC curve's range is 0-1 on both axis, the range of the area will always be [0-1]. Note however that random guessing technique will produce the diagonal line between (0,0) and (1,1) which has an area of 0.5. Thus no realistic classifier should have an AUC less than 0.5 making the new range [0.5-1].

6.3 Approach & Methodology

The objective is to generate receiver operating characteristic (ROC) curves showing probability of detection vs. false alarm rate while varying threshold for the four different RFI baseline

detection algorithms. This will facilitate in evaluating the performance against different RFI sources through computing the area under the ROC curves to compare performance.

6.3.1 General Approach

The SMAP radiometer engineering test unit (ETU) was designed to verify system performance, including RFI detection and mitigation. Artificial injected RFI signals were primarily of the pulsed-sinusoidal type and the continuous wave type. A variety of pulse widths (for the PW), amplitudes and frequencies were utilized with the goal of spanning low (~0.1K) to high-level (~1000K). The basic RFI scanning plan was to have a sequence of ~30 seconds RFI-on followed by another sequence of ~30 seconds off. The probability of detection is to be calculated by detecting when the RFI source is turned on, retrieve the number of samples that is flagged while varying beta, then divide the number by the total number of samples. For simplicity and approximation reasons, the FAR is to be calculated when the RFI source is off, retrieve the number of samples that is flagged while varying beta, then divide the number by the total number of samples. Each 30 second period correspond to a number of samples n depending on the integration period as well as the domain of operation (time or frequency). Figure 20 shows the big picture process of the general coding approach to each sample detector. Retrieving the number of samples per threshold will become necessary when generating the ROC curves.

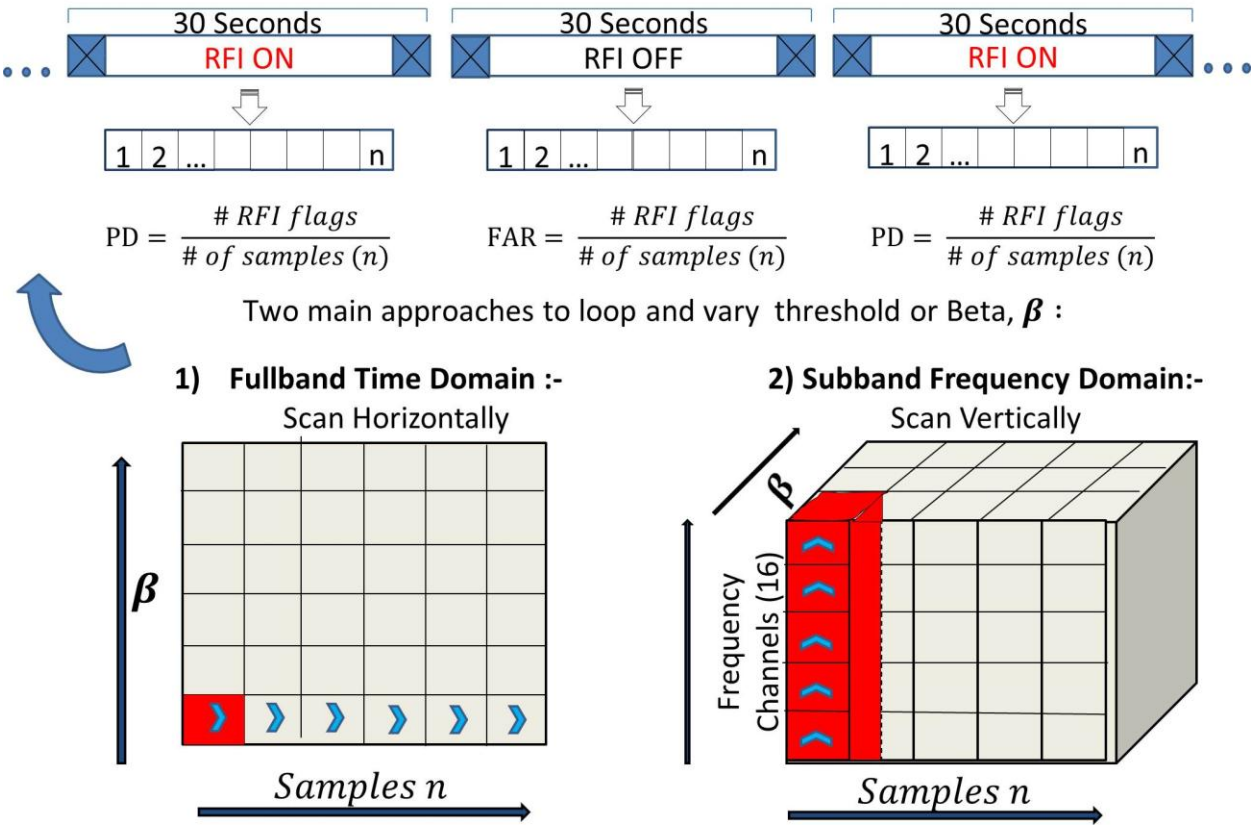


Figure 20 : Coding Method & Logic

Notice that the dimensions differ when facing fullband time domain versus subband frequency domain.

First, let's look at the fullband domain data. Referring back to pulsed-detection and time-domain approaches, a flagged sample would become sufficient for only considering that sample. Thus for each given threshold, a number of flagged samples will be retrieved. Varying the threshold, will create a simple two-dimensional table of values {threshold (β), number of flags}. From there it becomes a matter of dividing the number of flags with the number of samples to attain either the PD or FAR depending on the whether RFI is ON or OFF giving a final product of {β, FAR} or {β, PD}. To produce the ROC curves, equating the set of table and graphing them against each other would generate {FAR, PD} since the threshold values conducting for both logics are set the same.

On the other hand, subband domain can become three-dimensional due to the addition of frequency channels spectrum to the picture. Each sample n, has 16 frequency channels. Thus if any particular n is flagged, then the associated frequency spectrum channels must be flagged as well. This makes sure that in time-domain, where actual measurements are conducted, no corrupted signal at a specific time is accounted for. Managing to attach all frequencies to a single sample at a given threshold, β, will retrieve the same final result as fullband. The final result will still contain {FAR, PD} while varying threshold levels.

6.3.2 ROC Processing Code

The logic behind all different detectors and program processing language will be presented for clarification of computations. Refer to Chapter 5 for a background of each detector. Since the PD and FAR processes are very similar in computation, the same code was used for attaining both. The input and output domains were the only different parameters in the operation. Distinguishing between when the RFI was present or not is necessary to determine the input and thereby categorize the output. Data sets of Level 1A and Level 1B products were stored in a directory of inputs. Boundary files are stated as the files where switching between RFI ON and OFF occurs. Often, these files are neglected from calculations to avoid biasing different range files into one analysis. Various RFI baseline detectors will use products accordingly. Below is the documentation of the all the ROC processing code.

A. Time-Domain Code Logic

Time-Domain Equation: $\boldsymbol{T_A(t) - m(t) \geq \beta_{td}\sigma_{td}(t)}$

Objective: The time domain function generates the PD or FAR for the "time domain" RFI detector given the input files index number and the number of scans after the index file.

Inputs: Level 1B Product

i) Generate a loop that is able to compound time domain temperature samples from several different non-boundary files. These samples will be the basis of our data examination. Labeled in the equation "$T_A(t)$".

ii) Sort data to get lowest 90% of samples to compute the mean, "$m(t)$" and standard deviation, $\boldsymbol{\sigma_t}(t)$ of them.

iii) Set up a "for" loop around the time-domain function logical statement to vary the threshold beta, labeled $\boldsymbol{\beta_{td}}$ in the equation above.

iv) Create a 3D matrix of 1x number of samples x beta threshold versus number of samples to evaluate performance scaled from [0 (i.e. detecting nothing, large beta) to 1(i.e detecting everything, small beta)].

v) For each beta, evaluate the equation above to retrieve whether the logical statement is high versus the number of samples.

vi) Retrieve the number of times it was flagged vs. the threshold value into a 2D matrix, and divide by the total number of samples to get the PD or FAR.

B. Cross-Frequency Code Logic

Cross-Frequency Equation: $\boldsymbol{T_A(f) - m(f) \geq \beta_{cf}\sigma_{cf}(f)}$

Objective: This function generates the PD or FAR for the "Cross-Frequency" RFI detector given the input file index number.

Input: Level 1B Product

Note: Cross-Frequency is similar to time-domain baseline detection in terms of logical computation. The only difference is that cross-frequency looks at samples in the frequency

domain where there are 16 subbands or channels. Thus only significant changes will be mentioned below.

i) Sort data to lowest 12 frequency subbands (N=4). Removing the highest 4 frequency channels will allow consistency in computing mean and standard deviation.

ii) Create a 4D matrix with the additional dimension being the 16 channels corresponding to each time domain.

iii) Check for every given sample, for each time sample and threshold beta, if any of the subbands (1-16) are 1; flag the corresponding time domain value.

A. *Kurtosis Detection Code Logic*

Kurtosis Equation: $|\boldsymbol{K} - \boldsymbol{K_{nom}}| > \boldsymbol{\beta_k \sigma_k}$

Objective: This function generates the PD or FAR for the "Kurtosis" RFI detector given the input file index number and the number of scans. The code logic will operate on both fullband and subband data.

Inputs: Level 1A Product

i) Generate a loop to generate compounded values of all four moments of fullband and subband data. The loop scanning number will be retrieved from the main automation function.

ii) For the fullband, start varying threshold beta starting at a very low value of beta to ensure that PD and FAR values reach 1.

iii) Call the kurtosis detection function that generates RFI flags for a given data sample inside the threshold varying for loop.

iv) Make sure to remove any unwanted, unidentified values labeled “NaN”.

v) Find flagged columns for both vertical and horizontal polarization and average the final number of flagged samples.

vi) Generate a 2D matrix with each of the beta threshold values along with their corresponding number of flagged samples.

vii) Compute PD or FAR by dividing the 2D matrix created by the total number of samples.

viii) Extract the PD or FAR with account to step sizing that was generated due to starting beta at a very low value.

ix) Repeat steps for the subband data, and follow the same logic of Cross-Frequency in sampling in frequency domain.

B. *T3 & T3 Detection Code Logic*

Polarization Equation: $|\boldsymbol{T_{3,4}}| \geq \boldsymbol{\beta_{3,4}\sigma_{3,4}}$

Objective: These functions generate the PD & FAR for the "Third & Fourth Stokes" Fullband & Subband RFI detector given the input file index number and the number of scans. Performance will be evaluation over both fullband and subband products.

Input: Level 1B Product

i) Generate a loop to compound third and fourth stokes temperature samples for fullband and subband data after removing “NaN” values. The loop scanning number will be retrieved from the main automation function.

ii) Use the T3T4 full and subband function to retrieve RFI flags while varying threshold Beta. Logic in retrieving fullband and subband samples are identical to functions above.

E. “RFI_Information Function”:

This function saves and generates RFI source information. The code was attained and modified from Dr. Joel Johnson's code.

F. "*ROC_Automation Code Logic*":

This code processes through a given directory to generate PD **and** FAR for all four different detectors in both fullband and subband. First part of code consists of Dr. Joel Johnson's code. For reference see"process_integrate3_flight" file. Second part consists of code designed to read through a directory using information retrieved from previous code about the "Goodfiles" to scan through.

Objective: Objective is to automatically scan through a directory and generate FAR and PD quantities for all detectors in use.

Inputs: "RFI_Information", "time_domain_ROC", "cross_frequency_ROC", "kurtosis_ROC", "T3T4_ROC", and "T3T4_16_ROC".

i) Dr. Joel Johnsons' "process_integrate3_flight" creates a matrix of nonboundary files with the index of each of those files labeled. They are named "goodfiles". The length of that matrix is how many non-boundary files were in that directory.
ii) The sequence of files follow the RFI_ON, RFI_OFF logic. A sequence of files either with RFI source turned on or off are listed, followed by a gap of boundary files until the next scan is yet to come.
iii) These sequences usually tend to be either 4 or 5 files long. The sequences are usually separated by 3 files. However, these numbers aren't constant. Thus, a detection algorithm must be created to detect how long each sequence is as well as when the following sequence will start.
iv) After examining some of the goodfiles outputs, I noticed short sequences usually 1 or 2 files were classified as "good". Thus, for a conservative approach, I made sure that the first file in

the directory is followed by at least three files of the same sequence and named the new directory “GF”.

v) To determine the number of scans and the starting transition files I started by taking the difference of GF. This will either return 1, meaning the files are in a sequence, or more than 1 meaning they are in transition.

vi) I found where the difference was more than one in the matrix and named that the transition file. Also, adding 1 to that matrix to classify that first file as a transition file as well.

vii) Finally, the number of files or scans of each sequence are computed by determining the difference of those transition files.

viii) Since we need both PD and FAR values to generate the ROC curves, each loop incident must take increment by 2 files. This will ensure that one of each RFI_ON and RFI_OFF are included in the matrix. The “for” loop will start from 1, incrementing by 2, to the length of the GF matrix.

ix) Compute the first of two index files by calling the specific loop index of transition matrix that was created. The second transition file will be the very following index in the transition matrix.

x) The number of scans will be retrieved by calling the specific loop index of the scan matrix that was created earlier. The following scanning number will be the same index plus 1.

xi) Check whether RFI source is turned on or off for the first file index to compute PD and FAR accordingly.

xii) Call all RFI detector functions with inputs of the index file numbers as well as the number of scans for each sequence.

6.4 Results & Discussions

The performance of all four detector algorithms were analyzed and compared relative to each other. To resemble fair comparisons, detection algorithms were tested against the same sources of RFI. They were tested against two main waveforms, the pulsed RFI and continuous wave RFI. Figure () plots ROC curves for all detectors against a “9K” powered continuous wave RFI. Since the cross-frequency algorithm uses only frequency resolution and no time resolution, its performance should be best for RFI that is localized in frequency. Since continuous wave (CW) RFI are localized in frequency by definition [5], as expected, the cross-frequency detector performed

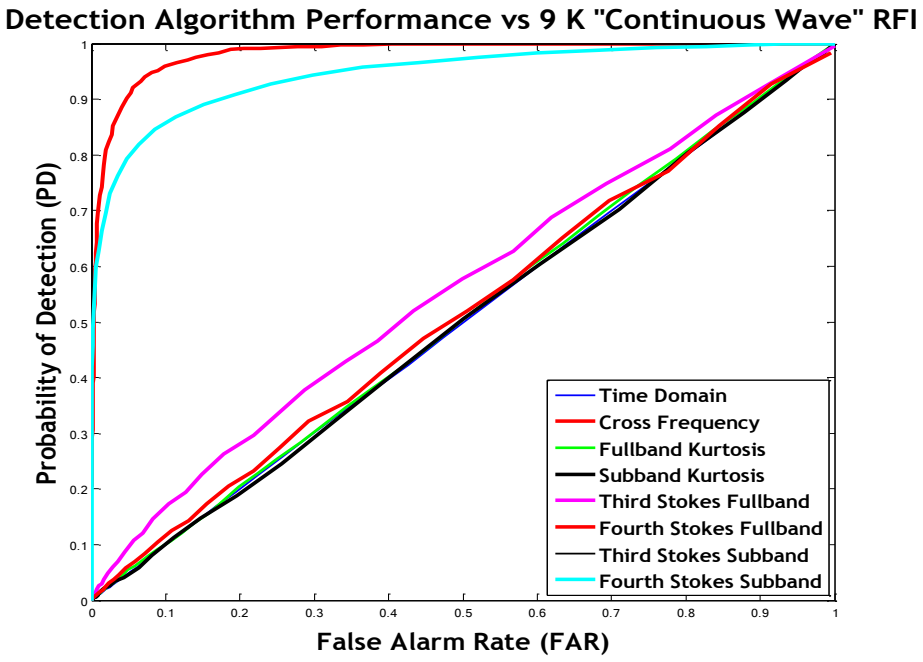


Figure 21 : ROC Curves for CW RFI

very well against CW RFI relative to other detectors as seen in Figure 21.

On the other hand, the same examination was conducted against pulsed wave RFI as seen in Figure 22. The cross frequency detection also seemed to outperform other detectors at the given

instant. Note however, that even though cross-frequency performed better than most other detectors at these two types of RFIs, cross-frequency doesn't necessarily perform better at all times.

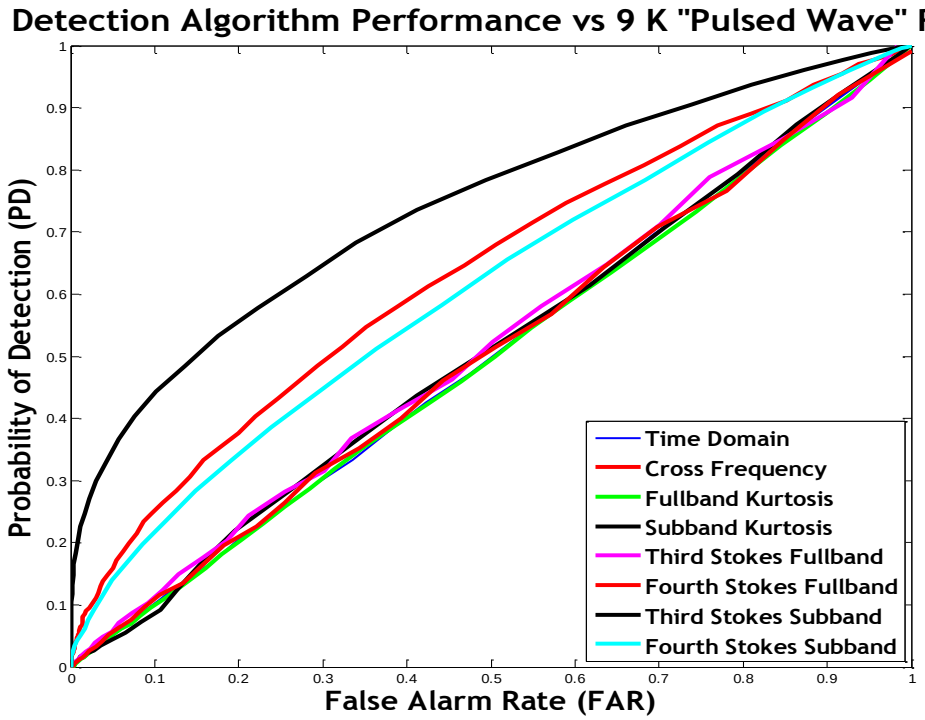


Figure 22 : ROC Curves for PW RFI

Polarization detectors outputted inconsistent behavior towards same RFI detectors. The reason behind this is still under investigation, however, the third and fourth stokes algorithms could be neglected for this particular examination. Realistically these data sets were conducted so that the horizontal and vertical polarizations are identical. Thus the cross correlation of these outputs should always near zero.

There were multiple RFI types conducted with different frequencies, amplitude, and pulse properties. The general performances of these detectors were a higher performance at frequencies that are very close to the operating bandwidth of [1400 -1427]. This reflects on the

fact that typical RFI sources will have a source frequency that is very similar to that of the SMAP satellite. Also, RFI sources with moderately high amplitudes showed the best performance by the detectors. This was realized that very high amplitudes caused disturbance and high variations in the data sets provided, making the detection process less efficient. RFI algorithms also showed best performance against low to moderate pulse widths.

7. CONCLUSION

The SMAP satellite is expected to launch in October 2014. The satellite radiometer satellite has been successfully shipped to the Jet Propulsion Laboratory (JPL). Continual studies on RFI algorithms and analysis will be performed as variations in space and flight tests will factor into the equation. The logic behind the detection algorithms will be implemented into the operating software of the satellite rather than theoretical testing in Matlab.

Properties of statistics were studies to understand the hardware components of the system. Such statistics included the first to four moments of mean, variance, skewness, and kurtosis. These statistical measures were able to reflect on how well the RDE was performing over time. Generally, for uniformly distributed Gaussian function, variations in skewness and kurtosis need to be at a minimum to ensure optimal performance. Analysis of the performance can help engineers building the RDE to debug certain errors encountered in testing.

Calibration was another concept introduced to transform preliminary RDE data into actual physical temperature brightness. Calibration is done with two data sets available of known measurements and ancillary information to understand the actual performance and measurements outputted by the RDE.

The RFI model presented into analysis of mitigation analysis was narrowed down to two main types: pulsed and CW. They represent the main sources of RFI from the previous studies conducted at the L-band and from previous similar missions encountered such as SMOS and Aquarius. RFI simulations have been performed for the two main types, pulsed and CW, to determine the algorithm performance of various detection methods. Other special cases of RFI were also included in the study to account for other random sources. Due to the difficulty of knowing what type of RFI will interfere with the system, a conservative approach must be taken to ensure to corrupt data measurements are taken.

All types of RFI detection performance had similar logic in common. They were ultimately a function of the threshold (that returns the highest PD versus lowest FAR), the nature of the RFI encountered (whether it is continuous wave or pulsed), and radiometer system parameters such as the number of frequency channels and the integration period. Further analysis into optimizing the four baseline detection algorithms can result in an overall more efficient system. Such analysis can include ROC and AUC computations of the different detection algorithms.

REFERENCES

[1] Piepmeier, J.R.; Johnson, J.T.; Mohammed, P.N.; Bradley, D.; Ruf, C.; Aksoy, M.; Garcia, R.; Hudson, D.; Miles, L.; Wong, M., "Radio-Frequency Interference Mitigation for the Soil Moisture Active Passive Microwave Radiometer," *Geoscience and Remote Sensing, IEEE Transactions on* , vol.PP, no.99, pp.1,15,

[2] D. Entekhabi, E.G. Njoku, P.E. O'Neill, et al. "The Soil Moisture Active Passive (SMAP) Mission," Proceedings of the IEEE, Vol. 98, No. 5, pp.
704-716, May 2010.

[3]Corbella, I., F. Torres, A. Camps, N. Duffo, M. Vall-llossera, K. Rautiainen, M. Martin-Neira, A. Colliander, "Analysis of correlation and total power radiometer front-ends using noise waves," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 11, pp 2452-2459, Nov. 2005.

[4] De Roo, R., S. Misra and C. S. Ruf, "Sensitivity of the Kurtosis Statistic as a Detector of Pulsed Sinusoidal RFI," *IEEE Trans. Geosci. Remote Sens*., vol. 45, no. 7, pp 1938-1946, July 2007.

[5] Güner, B., N, Niamsuwan and J. T. Johnson, "Performance study of a cross-frequency detection algorithm for pulsed sinusoidal RFI in microwave radiometry," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 7, pp 2899-2908, July 2010.

[6] C.D. Crandall, "Survey of Potential Radio Frequency Interference Sources," Space and Comm. Tech. Dir., Naval Research Laboratory,
Washington, DC, Memo. Rep. NRL-MR-4200, May 13, 1980 [Online]. Available: torpedo.nrl.navy.mil/tu/ps/doc.html?dsn=352081.

[7] L. Li, E.G. Njoku, E. Im, P.S. Chang and K. St. Germain, "A preliminary survey of radio-frequency interference over the U.S. in Aqua AMSR-E
data," IEEE Trans. Geosci. Remote Sensing, vol. 42, no. 2, pp. 380-390, Feb. 2004.

[8] E.G. Njoku, P. Ashcroft, T.K. Chan and L. Li, "Global survey and statistics of radio-frequency interference in AMSR-E land observations," IEEE
Trans. Geosci. Remote Sensing, vol. 43, no. 5, pp. 938-947, May 2005.

[9] L. Li, P.W. Gaiser, M.H. Bettenhausen, and W. Johnston, "WindSat Radio-Frequency Interference Signature and Its Identification Over Land
and Ocean," IEEE Trans. Geosci. Remote Sensing, vol. 44, no. 3, pp. 530-539, March 2006.

[10] S. Mecklenburg, M. Drusch, Y.H. Kerr, et al., "ESA's Soil Moisture and Ocean Salinity Mission: Mission Performance and Operations," IEEE
Trans. Geosci. Remote Sensing, vol. 50, no. 5, part 1, pp. 1354 – 1366, May 2012.

[11] Handbook of Frequency Allocations and Spectrum Protection for Scientific Uses, National Research Council of the National Academies,
Washington, DC: The National Academies Press, 2007.

[12] G. Lagerloef, F.R. Colomb, D. Le Vine, et al., "The Aquarius/SAC-D Mission: Designed to Meet the Salinity Remote-Sensing Challenge,"
Oceanography, vol. 21, no. 1, pages 68 – 81, March 2008.

[13] J.R. Piepmeier and F. Pellerano, "Mitigation of Terrestrial Radar Interference in L-Band Spaceborne Microwave Radiometers," Proc. IEEE Int.
Geosci. Remote Sensing Symp. (IGARSS), Denver, CO, pp. 2292-2296, July 30-Aug 4, 2006.

[14] S. Misra and C. S. Ruf, ``Detection of radio frequency interference for the Aquarius radiometer,'' IEEE Trans. Geosc. Rem. Sens., vol. 46, pp.
3123--3128, 2008.

[15] National Research Council of the National Academies, .*Spectrum Management For Science in the 21st Century*. 1st Ed. 1-226. Washington
D.C: The National Academies Press, 2010. Print.

[16] J. Park et al, “Airborne L-band RFI observations from the SMAPVEX08 campaign and associated flights,” *IEEE Trans. Geosci. Rem. Sens.*, vol. 49, pp. 3359-3370 , 2011.
[17] Fawcett, Tom. "An Introduction to ROC Analysis." *Science Direct*. (2005): 863-870. Print.
[18] For Mission site visit : http://smap.jpl.nasa.gov/